DOCUMENT RESUME

ED 021 778 24 SP 001 504

By- Hamreus, Dale G.
SELF INSTRUCTIONAL MATERIALS FOR RESEARCH TRAINING. SUPPORTING DOCUMENT TO A FINAL REPORT.
Oregon State System of Higher Education, Monmouth. Teaching Research Div.
Spons Agency- Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No- BR-7-1096
Pub Date Aug 67
Grant- OEG-1-7-071096-3873
Note- ~~336p.~~ 349p. $1.50
EDRS Price MF- ~~$1.25~~ HC-$13.52
Descriptors- BEHAVIORAL OBJECTIVES, DATA ANALYSIS, *EDUCATIONAL RESEARCH, INDIVIDUAL INSTRUCTION, *INSTRUCTIONAL DESIGN, *MANUALS, MEASUREMENT, RESEARCH DESIGN, RESEARCH PROPOSALS, *SELF HELP PROGRAMS, *SYSTEMS DEVELOPMENT, TEACHER DEVELOPED MATERIALS
Identifiers- *Consortium Research Development Projects, CORD

This self-instruction manual, developed for use in the 1967 National Research Training Institute for Participants in Consortium Research Development (CORD) projects, is designed to provide the user with basic information needed to (1) plan and produce an improved instructional system and (2) plan and conduct research related to instruction. It was used in conjunction with practice sessions and criterion testing. Bibliographies appear in 6 of the 10 sections. In "Orientation and Overview" Jack V. Elding outlines the first part of the individualized course in terms of an 11-step system design for instruction development. Step 1, "The Specification of Behavioral Objectives," is presented by Casper F. Paulson. "Objective Analysis of Instructional Specifications" by Paul A. Twelker encompasses the 4 planning steps of the design. H. Del Schalock deals with the "Construction of Performance Measures" step, and Dale G. Hamreus with "Instructional System Development" (the final 5 steps of production, try-out, analysis, modification, and recycling). The testing of the system through research constitutes the second part of the course: "Research Design" by John Gordon, Jr., "Data Analysis" by James H. Beaird, and "Proposal Writing" by Jack Crawford. Appendixes include lists of research fund sources and information on USOE research support. (JS)

BR-7-1096
PA-24

SUPPORTING DOCUMENT A

to

FINAL REPORT

Project No. 7-1096 '24
Grant No. OEG-1-7-071096-3873

NATIONAL RESEARCH TRAINING INSTITUTE FOR PARTICIPANTS IN CONSORTIUM RESEARCH DEVELOPMENT (CORD) PROJECTS

SELF INSTRUCTIONAL MATERIALS FOR RESEARCH TRAINING

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

teaching research

SP001504

OREGON STATE SYSTEM OF HIGHER EDUCATION

Office of Education
Bureau of Research

Self Instructional Materials
For Research Training

NATIONAL RESEARCH TRAINING INSTITUTE

For Participants in
Research Development
(CORD) Projects

August 1967

TEACHING RESEARCH
A Division of the Oregon State System
of Higher Education

I - ORIENTATION

Evaluation Sheet

We need to know how well the ideas and issues in this manual are communicated to you. You are the test audience for this material. To remove or strengthen the weak spots, to retain or improve the strong ones (if any), an account of your learning experience as you read these sections is crucial. Use this sheet as you study and jot down your reactions. Don't be concerned with typographical errors. We're concerned with how the message is coming through.

SECTION #________ TITLE______________________________

Give topic, paragraph or sentence, page	Importance Rate 1 2 3 4 5 Low - High	Clarity Rate 1 2 3 4 5 Low - High	Suggested Improvements

Over-all Rating of the Section: Use a 5-point Scale

Important					Understandable				
1	2	3	4	5	1	2	3	4	5
Very Low				Very High	Very Low				Very High

(Circle one) (Circle one)

PLEASE PUT ANY ADDITIONAL COMMENTS ON BACK

SECTION I

Orientation and Overview

Jack V. Edling

The purpose of this manual is to communicate concepts which normally are presented by lecture. It contains the information you are responsible for in the course. But information can be known at different degrees of proficiency. Therefore, after you have learned the contents of the manual you will apply the concepts, where relevant, to various problems. These problems will be presented in "practice sessions" scheduled daily. Through this routine, you will become proficient with the concepts, using and manipulating them productively.

However, only a limited time is available to cover knowledge that others have learned over many years of experience. It may not be possible, therefore, to gain the complete mastery that would be desirable. After you complete the practice session, you will take a new type of test (called a criterion test) to determine your mastery of the concepts. If you demonstrate mastery, you will apply the concepts to your field and also get an evaluation of your creative work. If you do not exhibit mastery on the criterion test you will enter either individual or small group instruction. If time runs out you will be directed to bibliographical references to continue your studies.

You will probably have questions about various phases of this program; for example, why was the Teaching Research Division selected to prepare and conduct this specialized research training institute; or what is the nature of the concepts being taught; how were they selected; what is my schedule as the institute is conducted? These and related questions will be answered in the orientation and overview.

Orientation

Teaching Research Division was organized by the Oregon State Board of Higher Education in 1960. Its objective is to improve the effectiveness and efficiency of instruction without reference to grade level or subject matter area. One area of concern is the improvement of elementary and secondary education primarily through teacher education programs. Also, it is concerned with

industrial and military training in cooperation with the Division of Continuing Education (established by the same board). But, its primary mission is to cooperate with the institutions in the Oregon State System of Higher Education, and higher education institutions over the country, to improve the quality of undergraduate and graduate instruction.

In its seven years of existence the Division has completed more than 80 research projects. The staff has grown from 2 to more than 50 highly qualified researchers in the behavioral sciences or support specialists essential for the continuing work of the Division. It has received more than 40 grants totaling approximately two million dollars. Each grant was obtained after the preparation of a proposal and a critical review by experts advising either federal or privately supported fund granting agencies. The operating budget for the current fiscal year is approximately three quarters of a million dollars.

The division is engaged in three major classes of activity, each called a unit. The basic research unit is concerned with finding limitations in existing theories and practices. Although this aspect is negative in character, the basic research unit seeks specific evidence of what is wrong in existing theories and practices and under what conditions. It then pushes for solutions to the gaps, the contradictions, and the limitations in existing knowledge. The contextual research unit, more positive in its approach, is applying what has been learned from research to the improvement of instruction. Its concern is the development of more effective and efficient instructional materials, tests, procedures and systems, combining them into the best possible learning experience for learners of all types.

For the first six years of the Division's history, a report or journal article was prepared on completed projects and placed on a library shelf to gather dust except for use by occasional scholars. But a year ago, the Directorate decided that more active effort should be applied to disseminate the work of the Division. A dissemination unit was organized to conduct institutes and employ other media to disseminate the findings of the Division.

In the spring of 1967, the Division applied to the U.S. Commissioner of Education for a grant to conduct a national research training institute for participants in Consortium Research Development (CORD) Projects. A separate proposal was submitted to the U.S. Commissioner of Education for a project to

develop a program of materials for a short term educational research training program. This manual is one result of that project.

Both the Teaching Research Division and the U.S. Office of Education believe that this Research Training Program will be relevant to requirements of institutions of higher education participating in CORD projects. The Teaching Research Division has as its basic objective (and its basic problem) production of change with minimum involvement. Minimum involvement is essential because if Teaching Research were to continue to assume the major responsibility for the development of new instructional systems it would be very limited in the number of departments or agencies it could serve. A more efficient way must be found to bring about change with less involvement by research and development specialists. The present attempt must be considered minimum involvement.

Our intention is to spend a concentrated two weeks attempting to train you in our techniques with no opportunity for continuing assistance. This may be below a realistic involvement necessary to create change on your campuses. However, through the Consortium Program, you are combining your efforts in attempts to affect necessary changes. Whether this combination will be adequate is unknown, for this effort is an experiment. If you significantly improve your courses, then you will answer the question as to whether the present approach is sufficient.

Overview

The ultimate objective of the Research Training Program is to provide you with basic skills necessary to plan and produce an improved instructional system and to plan and conduct research related to instruction. In order to achieve that ultimate objective it has been subdivided into seven enabling objectives each of which contains concepts and skills essential to attainment of the ultimate objective.

Through experience it has been learned that when an individual or department is given the charge, and the opportunity, to improve its instructional procedures it normally begins by doing more of what it is already doing. Most professors believe they already know what they need to do to improve their courses, and all they need is sufficient time and resources to do what they have always wanted to do--and an approximation to the perfect course will result. Unfortunately, this approach has been tried

extensively with the result that students are given the same content and the same ideas merely in a more elegant form. And the tragedy of this is, that empirical evidence has consistently shown that the more elegant form, and the more concentrated and sophisticated nature of the presentation, results in <u>less</u> learning by students. This is hard to accept. But, it is a consistent finding, and it has a rather firm psychological basis. The fact is that the requirements of the sophisticated learner, and what satisfies him, are very different from the requirements of the naive learner, and what satisfies him. The evidence indicates that the more sophisticated and informed the scholar, the less qualified he is to know the requirements of the naive learner!

Therefore, to merely afford a scholar more opportunity to prepare a course of instruction is not a sufficient condition, in and of itself, to insure the improvement of instruction. The steps in a process which has been demonstrated to be an effective means for developing more effective learning experiences are outlined in Figure 1.

The first task in the development of an improved instructional system is to identify criteria which will be used to determine when an instructional objective has been attained. If there is no identification, in measurable terms, of what the learner is supposed to learn as a result of exposure to the instruction, then there is absolutely no basis for any later determination as to whether the new instructional system is any better than the one it is designed to replace. This first step (1) is to "Identify the Behavioral Objectives" of instruction. They are called "behavioral objectives" because they must be stated in terms which are capable of objective measurement. Without the reliability of objective measurement there would be little purpose in going to the trouble of collecting data aimed eventually at making a decision as to whether there was in fact genuine progress toward building an improved instructional system.

Of course, most professors believe that what they teach, and what is important, is not subject to this kind of objective assessment. This issue, and the skill involved in identifying behavioral objectives, is the topic of the section prepared by Dr. C. F. Paulson. Dr. Paulson has been conducting research on methods for teaching the writing of behavioral objectives and has been actively developing instructional systems for six years. He is thoroughly familiar with the issues. He is considered as one of the national leaders in this rapidly developing innovation.

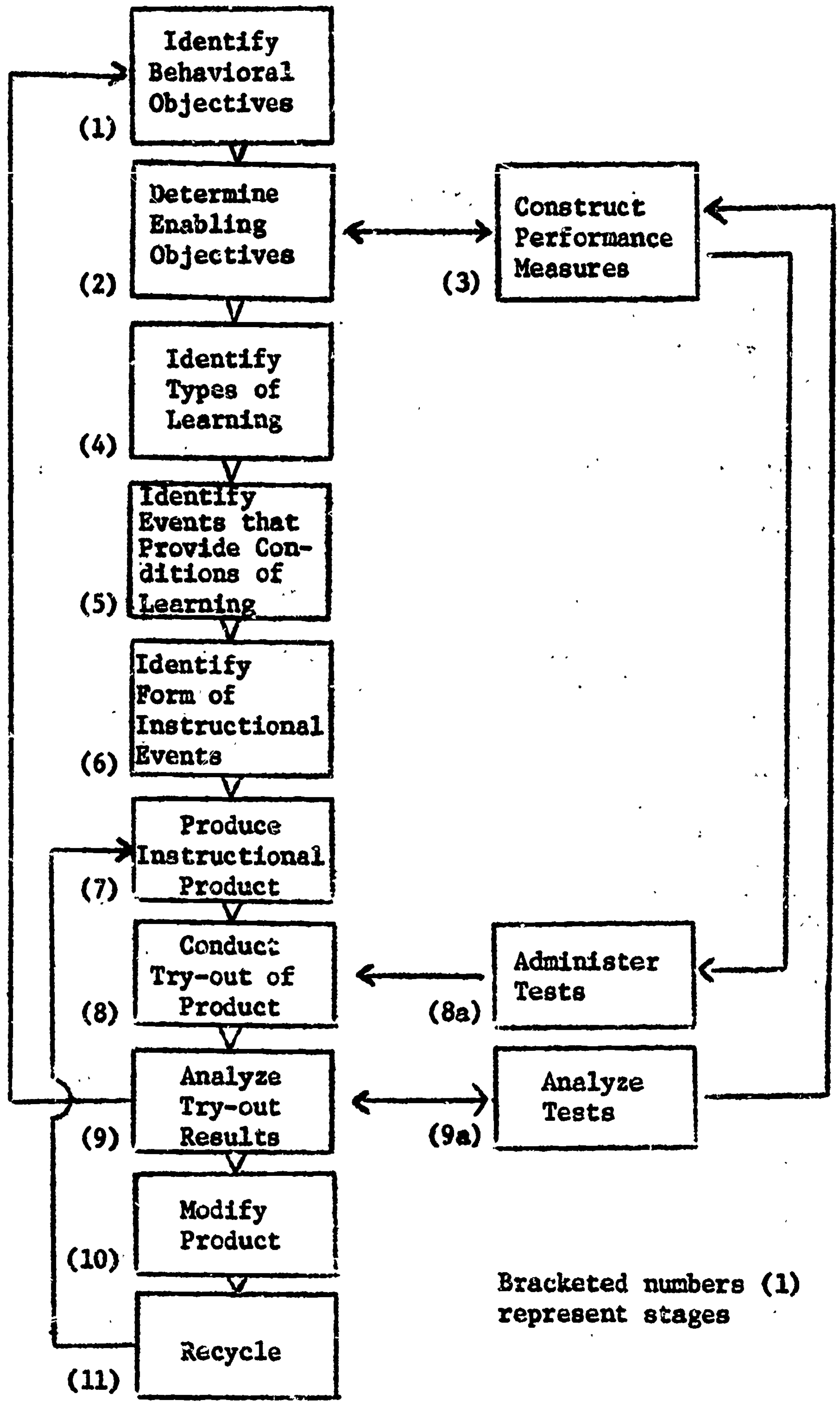


Figure 1. Major Stages in System Design for Developing Validated Instructional Systems

The prevalent approach to stating objectives might be classified as an information approach, i.e. objectives are stated in terms that boradly outline a content area. The learner might be expected to "know" information on any topic from Aachen (a German city) to zymase (yeast) and the professor would, through some manner, determine the adequacy of this information about the topic and, eventually, rate the learner's performance. The difficulty with this approach is that for all practical purposes there is an infinite amount of information that can be supplied on any object (and the quantity of this information is increasing hourly). A more rational and effective approach must be developed to determine what is to be learned or educators will be faced with the prospect of having to teach more than can be learned in a lifetime—on every single topic. This impossible situation can be resolved through the identification of behavioral objectives, and at the same time it affords the basis for significant improvements in learning effectiveness.

Once the ultimate objective has been identified there is a requirement then to determine the sub-behaviors (or elements) involved in that ultimate behavior. The analysis of any complex behavior reveals that it is composed of less complex behavioral elements which are essential to the effective and consistent performance of the more complex behavior. This second step (2) is shown in Figure 1 as "Determine Enabling Objectives." For example, the task of driving a car (as an ultimate objective) involves knowing how to start it, how to stop it, how to steer it, rules of the road and many other "elements." Obviously, each one of these sub-tasks or elements can become a training objective in and of itself. Therefore, it too is an objective, but it is merely an objective in the sense of enabling the more ultimate behavior to be performed.

Figure 1 illustrates that from the analysis of the ultimate objective (Step 2), two other kinds of activities are generated. One of these activities (Step 3) indicates that identification of enabling objectives permits the construction of performance measures. While this is perfectly true, for the time being we think it may be useful to set this activity aside and follow down the vertical line of blocks to learn the other kinds of activities generated from determining enabling objectives. Step four (4) indicates that when one has determined the enabling objectives it is possible to classify these into various types of learning. One type of learning might involve simple association, such as calling an object by a name, while a much more complex type of learning might involve decision-making or problem-solving where

objects must be classified, evaluated, and manipulated in very complex ways.

Depending upon the kind of learning that is involved, it is possible to identify the kinds of events that provide the optimum conditions for learning each kind of behavior (Step 5). Once the events have been identified that will permit optimal learning, then it is possible to further identify the form of those instructional events and specify them in a manner detailed enough to permit the actual construction of an instructional system. This process is somewhat analogous to that followed in the building industry. Step 1 may be likened to the basic decision to build a house, a garage, or an office building. Once the ultimate purpose of the building has been established, it then becomes possible to analyze more specifically what is desired. Of course, no analogy can be perfectly comparable but the general task of determining the various kinds of spaces desired and their approximate sizes or proportions is similar to the task of determining the enabling objectives. It too is an identification of the major elements which will comprise the total structure. Carrying the building analogy through the next three steps (4, 5, and 6), it can be seen that the bedrooms, for example, will have different kinds of requirements for the different kinds of people they are to house. When these requirements have been clarified, it will be possible to specify with considerable detail their size and location. And finally, with all these decisions having been made, it will be possible to construct a detailed blueprint which would specify in a very detailed manner exactly the form that each room must take in the finished product.

This four-step process (2, 4, 5, 6) has been given the name of "objective analysis and instructional specification." This topic has been developed by Dr. Paul Twelker in Section III. Dr. Twelker is an educational psychologist who has been the recipient of several research grants in the area of programed instruction. He has developed instructional systems and is particularly adept in analysis and planning learning experiences designed to produce specified types of behavioral outcomes.

There are several complex issues involved in these operations. Many professors believe it is impossible to efficiently pre-plan the kind of experience that will be maximally effective with a given class on a given day. Some feel it is essential to "develop the subject as it moves along." There is some very good evidence on this issue and there is much to be said for the

flexible point of view, but it is not in opposition to detailed analysis and planning. Being flexible is most effective as a condition that comes after detailed analysis and planning--if optimal learning is to result. In other words, the analyses and the procedures taught by Dr. Twelker are useful and necessary in the development of new instructional systems, regardless of how the instructor might modify or supplement that system when it is finally used under optimal circumstances.

From the "blueprint" or the detailed "instructional specifications," it is possible to produce an instructional product or prototype for each enabling objective (Step 7). This version is then tried out, or tested, with students (Step 8) to determine whether, in fact, the intended learner can exhibit the desired behavior. Whether the learner can or can not perform the behavior identified in the enabling objectives, has some very complex relationships to the behavioral objectives originally identified. In some instances the analysis (Step 9) will reveal that the instructional specifications were in error. Under other circumstances, they will reveal that the instructional product is deficient. Under still other circumstances, the analysis will reveal that both specifications and the product are proper, or that they are both deficient. Depending upon the results of this analysis, the product is appropriately modified (Step 10) and recycled (Step 11) through the production and testing of additional materials, procedures, or systems until a product is designed that does enable learners to perform the desired behavior in an effective and consistent manner.

Steps 7 through 11 have been given the title of "Instructional Systems Development." Dr. Dale Hamreus develops this topic in considerable detail in Section IV. Dr. Hamreus has had extensive experience in developing instructional systems, and over a four-year period directed a major grant from the Ford Foundation involving approximately fifty instructional systems ranging from the elementary school through the college level in most of the academic disciplines--including the physical and social sciences, humanities, mathematics and even more generalized and interdisciplinary subjects such as "use of the library" and "study skills." The insights which he has gained over these years of experience hopefully will be useful in developing instructional systems that are valid, i.e., teach the behavior they were designed to teach.

Although the construction of performance measures was alluded to earlier, and is indicated as Step 3, discussion has been purposefully delayed to proceed through the steps in the developmental

process in the most direct manner. Figure 1 clearly indicates, however, that the performance measures are constructed very early in the developmental process and the two-way arrows between Stages 2 and 3 indicate that the construction of performance measures actually aids in determining enabling objectives. This is true because in analyzing the behavior to be measured, there is an analysis of the ultimate behavioral objectives. Further, by constructing performance measures early, the specific measures used to determine the quality of the ultimate behavior exhibited by the learner are identified, and these specifics are avoided in the instructional experience. Thus, an individual does not learn the answer to specific questions, or the solution to a particular kind of problem, in a rote way, but rather is required to solve a problem, or arrive at a generalization, or perform any other task in a manner which assures that he can perform the desired behavior in a meaningful context. Figure 1 also indicates that the measurement devices are useful in the tryout of the product, and the analysis of results aids in refining the measurement instrument itself.

When all steps have been completed, it can readily be seen that the instructional system is complete including all necessary performance measures. Even the term "performance measures" is intended to communicate that this does not necessarily refer to paper-and-pencil tests, but may include any conceivable way to bring an objective assessment to the behavior being taught.

The construction of performance measures is covered in Section V under the title of "Measurement." This section was prepared by Dr. Del Schalock. Dr. Schalock has spent the last five years developing new kinds of measures of complex behavior. Most of his professional career has been dedicated to the search for improved measures of the complex forms of human interaction often stated in our objectives, but somehow never quite measured in existing paper-and-pencil tests. The film-tests" which he has designed constitute one of the newest and most promising approaches to one of the most difficult problems in measurement.

At this stage, of course, the process of development has been completed. If one had complete mastery of the concepts and the skills described above, it would be possible to build a complete instructional system that would be valid and would achieve the ultimate behavioral objectives originally identified. Unfortunately, the process is not that simple. When the time comes to adopt any instructional system, the varying characteristics of learners, varying instructional situations, teachers

that employ materials in varying ways, the varying context in which materials are used, etc., all contribute to differences in the effects that a system may have. In order to test the system in these varying contexts, and to determine its effectiveness when employed in different ways, it is necessary to establish a "research design" that establishes the conditions for making valid inferences.

Dr. Jack Gordon explains the principal considerations for research design in Section VI. Dr. Gordon received his Ph.D. in Research Design and Analysis at Michigan State University. His professional work has concentrated on problems related to research methodology, and he is nationally recognized and consulted as an expert in this work. He has sought to prepare materials which will help you critically analyze proposed research design and thus eliminate the sources of invalidity in the conditions of an experiment. Hopefully, this training will enable you to critically analyze your own designs and to perfect them in order that you will be enabled to make valid inferences about the effectiveness of instructional systems which you produce in various contexts of use.

Sections VII and VIII deal with analysis of data gathered when the research design is applied. Many professors whose basic discipline is not in the mathematical area hesitate to become involved in these analyses because they believe it is beyond their competence. Dr. James Beaird, who has prepared Sections VII and VIII, has such a command of these techniques that he believes the basic concepts and basic decisions as to which kind of analysis is appropriate for given kinds of questions can be readily mastered by anyone who is willing to learn a few basic considerations. There is general consensus in this age of the computer that it is <u>not</u> essential to learn the actual skills of calculation. Such training for many professors may constitute a waste of much human effort. However, the basic skill necessary to specify the kinds of information desired are essential for the person who wants to design and conduct research on instruction. Dr. Beaird has devised a unique approach to assisting the novice in data analysis to enable him to give the kinds of direction necessary to appropriate experts in order that even a novice may attain the kind of information required to draw the valid inferences alluded to above.

Finally, it is necessary to possess skills to obtain support for both developmental and research work and to report on the effectiveness of such systems once they have been developed. These activities are described in Section IX under the topic of

"Proposal Preparation." This section has been prepared by Dr. Jack Crawford. Dr. Jack Crawford is a clinical and experimental psychologist who has had extensive experience in educational and psychological research and has received numerous research grants. His extensive experience in both conducting and administering research has provided him some insights in "grantsmanship" which may be invaluable to you when you seek support for your work. Obviously, there are no pat answers, or formulae, to present an idea in such a way that you will be guaranteed support. However, there are some rules which must be followed and there are some practices to avoid. These suggestions or guidelines can help in making a presentation that is favorably reviewed.

During the next two weeks you will be deluged with massive amounts of information relevant to research and development. Perhaps one of the most useful things you can do to integrate this information is to try to organize it around some research idea or project of your own. It is suggested that you construct a brief outline for a research proposal (see Section IX) and as each topic is developed apply it to your own project. Check it daily with the staff member leading your small group discussion session, and when it has his endorsement, check it with Dr. Edling for any additional comment to get a second reaction. A proposal so developed is really one of the most useful outcomes of the institute. If you have developed a sound research proposal by the end of the institute you will have demonstrated attainment of the ultimate objective of the institute--and one of the significant enabling objectives of the CORD program.

The detailed official daily schedule of sessions follows:

Good Luck

APPENDIX A

Official Institute Schedule

Sunday August 20 — Registration and dormitory check-in.

	8:00-11:30 a.m.	1:00-4:30 p.m.*	7:00-10:00 p.m.
Monday August 21	Section I-Orientation (Dr. Edling)	Independent Study-Specify'g Objectives	Social Event
	Section II-Presession Evaluation (Dr. Crawford)	Orientation (Dr. Edling)	Social Event
Tuesday August 22	Section I-Practice Session-Specifying Objectives. (Dr. Paulson)	Small Grp. Session-Specify'g Objectives (Dr. Edling-Staff)	Independent Study-Objective Analysis
	Section II-Independ. Study-Spec. Obj's.	Practice Session-Specify'g Objectives (Dr. Paulson)	Independent Study-Objective Analysis
Wednesday August 23	Section I-Practice Sess.-Obj. Analysis (Dr. Twelker)	Small Grp. Session-Objective Analysis (Dr. Paulson-Staff)	Independent Study-Instruc. Syst. Dev.
	Section II-Small GpSs. Specify'g Objectives (Dr. Paulson-Staff)	Practice Session-Objective Analysis (Dr. Twelker)	Independent Study-Instruc. Syst. Dev.
Thursday August 24	Section I-Practice Sess.-Instr.Syst.Dev. (Dr. Hamreus)	Small Grp. Session-Instruc. Syst. Dev. (Dr. Twelker-Stf.)	Independent Study-Measurement
	Section II-Small Grp. Sess.-Obj. Analysis (Dr. Twelker-Staff)	Practice Session-Instruc. Syst. Dev. (Dr. Hamreus)	Independent Study-Measurement
Friday August 25	Section I-Practice Sess.-Measurement (Dr. Schalock)	Small Grp. Session-Measurement (Dr. Hamreus-Stf)	Independent Study-Research Design
	Section II-Small Grp. Sess.-Instr.Syst.Dev. (Dr. Hamreus-Staff)	Practice Session-Measurement (Dr. Schalock)	Independent Study-Research Design

*At 4:30 p.m. to 5:00 p.m. daily there will be a general discussion session in the auditorium for both Sections I and II to evaluate and clarify issues raised during the day.

Saturday
August 26

Sunday
August 27

Official Institute Schedule
(Second Week)

	8:00-11:30 a.m.	1:00-4:30 p.m.*	7:00-10:00 p.m.
Monday August 28	Section I-Practice Sess.-Res. Design (Dr. Gordon)	Small Grp. Session-Research Design (Dr. Schalock-Stf)	Independent Study-Data Analysis I
	Section II-Sml.Grp. Sess.-Measurement (Dr. Schalock-Stf.)	Practice Session-Research Design (Dr. Gordon)	Independent Study-Data Analysis I
Tuesday August 29	Section I-Practice Sess.-Data Anal. I (Dr. Beaird)	Small Grp. Session-Data Analysis I (Dr. Gordon-Stf.)	Independent Study-Data Analysis II
	Section II-Sml.Grp. Sess.-Res. Design (Dr. Gordon-Stf.)	Practice Session-Data Analysis I (Dr. Beaird)	Independent Study-Data Analysis II
Wednesday August 30	Section I-Practice Sess.-Data Anal. II (Dr. Beaird)	Small Grp. Session-Data Analysis II (Dr. Gordon-Stf.)	Ind. Study-Proposal & Report Writing
	Section II-Sml.Grp. Sess.-Data Anal. I (Dr. Gordon-Stf.)	Practice Session-Data Analysis II (Dr. Beaird)	Ind. Study-Proposal & Report Writing
Thursday August 31	Section I-Practice Sess.-Proposal & Report Writing (Dr. Crawford)	Sml.Grp.Sess.-Prop. & Report Writing (Dr. Beaird-Stf.)	Social Event
	Section II-Sml.Grp. Sess.-Data Anal.II (Dr. Beaird-Staff)	Prac.Sess.-Proposal & Report Writing (Dr. Crawford)	Social Event
Friday Sept. 1	Section I-Final Criterion Meas. (Dr. Crawford)	Summary & Evaluation of Institute-Dormitory Check-out (Dr. Edling)	
	Section II-Sml.Grp. Sess.-Proposal & Report Writing (Dr. Beaird-Stf.)	Summary & Evaluation of Institute-Dormitory Check-out (Dr. Edling)	

*At 4:30 p.m. to 5:00 p.m. daily there will be a general discussion session in the auditorium for both Sections I and II to evaluate and clarify issues raised during the day.

Participants may make travel departure schedules after 3:00 p.m. Friday. Transportation to Portland and Salem terminals will be provided.

II - OBJECTIVES

Evaluation Sheet

We need to know how well the ideas and issues in this manual are communicated to you. You are the test audience for this material. To remove or strengthen the weak spots, to retain or improve the strong ones (if any), an account of your learning experience as you read these sections is crucial. Use this sheet as you study and jot down your reactions. Don't be concerned with typographical errors. We're concerned with how the message is coming through.

SECTION #_______ TITLE________________________

Give topic, paragraph or sentence, page	Importance Rate 1 2 3 4 5 Low - High	Clarity Rate 1 2 3 4 5 Low - High	Suggested Improvements

Over-all rating of the Section: Use a 5-point Scale

Important						Understandable				
1 Very Low	2	3	4	5 Very High		1 Very Low	2	3	4	5 Very High

(Circle one) (Circle one)

PLEASE PUT ANY ADDITIONAL COMMENTS ON BACK

SECTION II

Specifying Behavioral Objectives

Casper F. Paulson

The Case Against "Specifying Objectives "

An interesting phenomenon occurs whenever educators prepare to document plans for instruction. Immediately prior to undertaking the business at hand, they prepare eloquent statements about values they hold dear, and believe they share with their patrons. This activity is variously described as "specifying objectives," "stating goals," or "describing aims." The regularity of the occurrence of this phenomenon is difficult to understand, because only infrequently do these statements appear to have any relevance to the subsequent planning activity. In most cases, the relationship appears to be analyagous to the playing of the national anthem before an athletic contest.

It is our purpose here to describe effective techniques for planning and developing instructional systems. It is not our purpose to perpetuate dogma. Thus it has seemed necessary to examine critically even such a generally accepted procedure as "specifying objectives." Careful examination of the effects and side-effects of the procedures generally subsumed under this label indicate that their positive contribution is usually negligible. Worse yet, they frequently involve behavior that is psychologically maladaptive, socially repugnant, and directly antithetical to the improvement of instruction.

Generally speaking, when someone expresses an intention to do something, to bake a cake, or build a home, or vacation in Hawaii, we take him seriously. We expect that he really will do what he says he intends to do. However, if a person continually expresses intentions to do certain things, but never does them, we soon learn to discount such talk as idle chatter. We come to believe that such a person has lost touch with reality, and that he achieves satisfaction from expressing dreams rather than achieving goals. Whether such behavior involves merely fantasy or autistic thinking, in other words whether he is merely dreaming and knows it, or he really believes that he can change his circumstances by merely wishing hard enough, we tend to feel sorry for such a person and regard him as a victim of himself, but harmless to others.

There may be serious consequences, however, when statements of intention are taken at their face value and subsequently prove invalid. If we hire a contractor to build a house, and later find that he lacks the competencies and resources to complete the job, the fact that he had good intentions does little to lessen our dissatisfaction. It is even possible that the expressed intentions were nothing less than fraudulent. Expressions of intent are like promissory notes. They are only as good as the ability and determination of their maker to fulfill them. The implication that many instructional objectives range from daydreams to autistic thinking is not accidental, nor is the implication that taking them seriously may have harmful consequences.

Unfortunately, the actual achievements of instruction are much more difficult to measure than those of other enterprises. The difficulty is compounded when objectives are expressed in terms that make them invulnerable to validation, as when they refer to events to occur in the distant future or in the inner recesses of the human mind or soul. We have been told that we will be able to separate true from false prophets by their fruits. It is difficult to believe in invisible fruit. It is hardly credible that seeds of instruction will lie dormant for an extended period of time, and then miraculously sprout, mature, flower, and bear fruit at some time in the distant future. It seems reasonable that true prophets would readily call attention to the fruits by which they could be identified.

Imagine, if you will, the effect on the progress of medical science if we could not differentiate surgeons from quacks, pharmacists from medicine men, antibiotics from sugar water. Imagine how much of our economic progress would have been possible if contractual agreements were entered into with such expansiveness, and honored with such prodigality, as is so commonly the case with educational objectives. If you can imagine these things, you can have some general idea of how much the statements of educational objectives have contributed to the progress of education to date.

The majority of educational objectives are fundamentally fraudulent. They have the capacity to deceive both those expressing them and those to whom they are expressed. They may be used both to camouflage our confusion and to propagandize the public.

To summarize the case against specifying instructional objectives, not only is this activity frequently an unproductive, even irrelevant, ritual; it frequently impairs intellectual effectiveness and impedes educational progress.

The Case for Specifying Instructional Objectives

Although the improper use of statements of instructional objectives has probably led to widespread skepticism and suspicion, on the one hand, and misplaced confidence on the other, the hard fact remains that any substantial and durable improvement in instruction would probably be impossible without them. Just the use of the term "improvement" implies that we will achieve something more fully, frequently, or efficiently, or something not previously achieved. When we have described this "something" and set out to achieve it, we have an "objective."

The only way we can demonstrate improvement is to show that we have achieved at least one of the above. Such improvements are quite unlikely to occur by chance. They will usually be the result of a concerted effort. Furthermore, if improvements are to survive, and lead to further improvements, they must not rely merely upon one person's opinion, but documented evidence that will persuade even the unconvinced. Mere opinion, even conviction, will not suffice. Learning achievements must be described in terms that are credible and significant to the users.

Improvements of instruction is synonymous with improvement in achieving instructional objectives. We can hardly improve our proficiency at achieving those objectives without improving our proficiency in defining them. While recognizing the multitude of pitfalls and abuses that are possible in stating objectives, the appropriate question is not whether, but how, we should state them.

What is an Objective?

At the very least, an objective must represent some point or event that is identifiable. You must be able to tell when you have arrived. If others are expected to pursue the same objective, they must likewise be able to determine when they have arrived. The number of potential objectives is, of course, infinite. They only become objectives when we set out to achieve them. But before we leave this point its importance should be emphasized. Just being able to recognize something of value can be of considerable importance.

We often read of cases where valuable stamps or coins or priceless antiques have been sold for ridiculously low sums because their owner did not appreciate their true value. Each of us, during our lifetime, has probably had pass through our possession a

number of coins that could have been sold for considerably more than their face value. It is quite possible that some person reading this has in his pocket one such coin. Yet only the trained observer, the informed numismatist, will be able to capitalize on such serendipitous occurrences. It is unlikely that a person could make a respectable living by passively waiting for such rare coins to pass through his possession. The professional soon learns to identify and pursue likely sources.

Thus we come to the next essential characteristic of an objective. There are an infinite number of potential objectives. They only become real objectives when someone sets out to reach them. This implies not only a committment to attain them, but some control over the conditions necessary for their attainment. One may wish for rain, but unless he can cause it to rain, rain is not a legitimate objective.

A true objective, then, should have these characteristics:

1. It is defined clearly enough that we can recognize it.

2. We can carry out whatever activities are necessary to make it occur.

3. We seriously intend to achieve it, even at considerable cost.

Many of the malfunctions of the objectives described previously can be attributed to the fact that they do not fulfill the above criteria. Thus, many objectives are either not identifiable, or we lack the means of achieving them, or we are not really committed to their achievement. Only when these three criteria are satisfied can we truly say that we have an objective.

What is a Specification?

When we build a house there are some decisions that we are willing to leave to the architect or builder, and some that we have made ourselves. When we supply a list of conditions, those things that must be included, the conditions that must be met, we have supplied a list of specifications. Many other conditions may be varied, but these conditions must be met. We must give certain specifications before the work can commence, but it is almost as

important for us to realize what we should and should not specify.

For example, if we tell an architect that we desire that the home be comfortable and attractive, we have not yet made specifications. He doesn't know exactly what we mean by these terms, and he will request some more specific guidelines. Our specifications must be explicit enough to allow him to draw up a detailed plan. It is important, however, that we specify the right things. If we are too specific about the materials to be used and the manner of construction, we may impose conditions that are too restrictive.

The most useful specifications that we could supply to an architect would probably be those that described how the house should function, rather than how it should be built. For example, we may wish to have entertainment zones sufficiently separated from the sleeping areas so that our children won't be disturbed when we have guests. We may have fondness for outdoor living and phobias against bathroom congestion. Supplied with such a list of functional objectives, and permitted sufficient latitude in the means to achieve them, an architect may be able to provide us with a much more effective plan than we would be able to draw up for ourselves.

Similarly, in constructing behavioral objectives it is important that we be specific, but it is also important that we be specific about the right things. We should be much more concerned about the valued outcomes than about the means to attain them.

Why "Behavioral" Objectives?

We can specify objectives in several ways. We can specify those learning activities that we intend to take place, or the instructional activities that we intend to perform. But these are actually only means to an end, which is that the students will "learn" something. The whole business of education is concerned with changing people in some way, with developing characteristics that were not there previously. Unfortunately, the changes that we seek to accomplish are usually difficult to observe. Yet instructional decisions, unless they are purely arbitrary, are based upon some kind of observations. Readily identified landmarks like topics covered, pages read, or time elapsed have a way of becoming pseudo-objectives.

To avoid this pitfall, it is important that we render learning achievements more conspicuous. There is nothing terribly novel

about this approach. Progress in virtually any science depends upon the ability to enhance the observability of phenomena of interest. The cloud chamber and the Geiger counter, litmus paper and microscopes, all serve this purpose. The real challenge in constructing behavioral objectives is not to focus interest on what is readily apparent (as some skeptics think), but to make more apparent that which is of most interest. This may involve structuring the students' environment so that their actions are more readily observed, or more carefully structuring out observations so that we are more sensitive to significant aspects of the students' behavior.

It requires more than a little skill to construct objectives in such a way that they will provide the instructor with the information that he needs to make instructional decisions effectively, without becoming involved in such extensive testing that all of the intended learning is stifled. A prudent gardener does not dig up his seed every morning to see if it is growing. But there is really not much difficulty in getting students to behave in an overt and observable manner, as most teachers will testify. The task is more one of structuring the situation so that behaviors of interest can be observed, and then determining whether such behaviors indicate the attainment of the instructional objective.

The Three Dimensions of an Instructional System

There are three crucial factors that will determine the ultimate impact that an instructional systems will have on the improvement of education. It suits our purpose here to think of these as "dimensions," and then to employ an analogy. All solid objects are considered to have three dimensions. Each of these dimensions is a factor determining the "size" of the object. If any of the dimensions is zero, the object simply doesn't exist. Similarly, an instructional system can be thought of as having three dimensions, each determining the "size" of its contribution to education. With the reader's forbearance, we will employ the mnemonic device of identifying these three "dimensions" as "The Three D's of an Instructional System: Design, Development, and Description." With respect to each of these dimensions, statements of objectives are of crucial importance.

Design. Whether we aspire to build the house of our dreams or an instructional system, there comes a time to translate our aspirations

into specific objectives. As anyone who has built a home can tell you, this is the time when the dreaming stops and the hard decisions begin. While it will very likely be necessary to scale down our aspirations until they are commensurate with our resources and appropriate to our circumstances, this does not mean that we abandon our dream in despair. On the contrary, it is the time when we should guard most tenaciously those parts of the dream that we valued most and surrender only those valued least. How well this can be accomplished will depend largely on the clarity and explicitness of our values.

Similarly, the planning of an instructional system should not signal the abandonment of philosophical goals, educational ideals, or esthetic values. In spite of all the unkind things that have been said earlier about intangible goals and lofty aspirations, they are not inherently evil. When it comes to stating objectives, they represent a very good place to start, but a very poor place to stop. The more dearly they are held, the more urgent it becomes to express them in a form that will allow us to plan effectively for their achievement. Such statements give us the general direction of the objective, but they do not describe an objective. It may be of interest to know that there is some excellent fishing east of here, but a rabid fisherman would want to know how to locate and identify some specific excellent fishing hole.

Just what design functions should a good objective serve? In addition to giving direction to the instructional task, it should delimit the task. It should describe something feasible of attainment with the resources at hand. A general goal such as appreciation of our cultural heritage not only requires transformation into more observable terms, but considerable delimitation if it is to be feasible of attainment. No one would argue that describing the historical origins of the Bill of Rights is all that is involved in an appreciation of our cultural heritage, but it does represent a legitimate objective that is highly relevant to the more general value statement.

Delimiting the objective allows the developer of an instructional system to get closure. One could never complete the task of developing an appreciation for our cultural heritage, but the point of attainment of a behavioral objective is readily identifiable. If the behavioral objective is to serve the purpose of providing closure and identifying completion of the instructional task, it must describe behaviors that will be observable within the spatial and temporal limitations of the instruction. Furthermore, its attainment should not depend upon external, uncontrollable events.

In addition to giving direction, focus, and reasonable limits to the design of an instructional system, the objective should be stated in a manner that facilitates the planning of appropriate instructional procedures to achieve it. We must have at our disposal the means of attaining an objective, or else it is no more than a wish. In this event, our instructional activity, like a rain-dance, functions to relieve our anxieties, rather than to achieve our objectives.

This is one reason why it is important to express the objective in behavioral terms. There is no scientific body of knowledge that speaks to the problem of developing hypothetical constructs. But educators and psychologists do know something about changing human behavior. It seems to be fashionable to deplore the dearth of scientific principles of instruction, but the difficulty perhaps lies not so much in our deficiency of knowledge as in our deficiency in stating instructional problems appropriately. It may be true that this young science is not yet possessed of ten talents, or even five, but this does not justify burying the talent that we do have.

Concurrent with the planning of the operations of an instructional system is the development of a test to determine whether the system achieves the stated behavioral objectives. This test differs from those used in conventional instruction in an important way: it is used to evaluate the instruction, not the student. Thus while the administration of a final examination usually signals the fulfillment of an instructor's responsibility to his students, the end of his task, the administration of the test usually signals not the end of the developmental task, but merely the inception of a new and crucial phase.

<u>Development</u>.

"The best-laid schemes o' mice an' men
gang aft a-gley."

The first trial of an instructional system will usually provide one with an opportunity to quote Robert Burns. It is rare indeed that even the most carefully designed instructional systems will function perfectly on the first trial. Even though well established learning principles may have been applied meticulously in the design stage, the implementation and combination of these often have unexpected or unwanted results. In a sense, the first draft of an instructional system is a kind of complex hypothesis to be tested.

The appropriate test of the system is not whether it was presented as planned, but whether it achieved as planned.

The function of behavioral objectives in this context is to provide a sensitive indicator of the performance of the system, and to inform the developer of the nature of its strengths and weaknesses. Good objectives enable the establishment of an efficient information feedback system that allows timely correction of the deficiencies of the system. It should be noted that the use of empirical procedures in developing an instructional system places a uniquely valuable resource person, the naive student, on the developmental team. Frequently instructional planners know too much, and have forgotten how difficult it is to learn. The student, by his learning performance, can show the instructor where too much previous knowledge had been assumed, where the instructional pace is too rapid or too slow, and where the desired learning is or is not taking place.

It is the system, not the student, that is being tested. It is fairly well known that a good student can learn in spite of poor instruction. What is not quite so generally appreciated is that effective instruction can overcome many learner deficiencies.

Description.

"If my net don't catch it, it ain't fish."

This homely saying is often directed at behavioral scientists, suggesting that they are preoccupied with overt behavior and totally unconcerned with anything not presently amenable to observation. No doubt many wondrous things slip through our nets, but that should not deter us either from fishing or developing better nets. But let us rephrase the saying:

"If my net hasn't caught it, I shouldn't try to sell it."

When the development of the instructional system has been completed and its effectiveness demonstrated, it becomes important to describe it in terms of its most relevant characteristics so that it may be used by others. If this is not done, the system constitutes an instructional "grab-bag."

The description should be intellectually honest. This requires, in the first place, that the system be described in terms of the instructional objectives actually attained and observed in a

a realistic trial. It is not appropriate or honest to describe "intended" outcomes as attributes of the system. While intentions may be attributes of the designer, only the observed achievement of these intentions is appropriately regarded as an attribute of the system.

Intellectual honesty is further served when we make clear our educational goals with explicit statements or operational definitions of just what we consider to be adequate evidence of the achievement of those goals. It is not sufficient to say that our students have developed an understanding of the democratic processes, even if our observations have convinced us that this is true. While most people would agree that such an objective is important, there may be wide disagreement about what behavior constitutes achievement of the objective. One instructor may consider that he has achieved the objective when his students have memorized the Constitution. Others may disagree wholeheartedly. The most honest and least deceptive approach is to describe the instructional system in terms of the observed attainment of behavioral objectives. If a potential user considers these objectives and this terminal behavior to be inappropriate, well and good. At least he has not been mislead by a deceptive label. It is up to him to determine, in his own case, the relevance of the instructional objectives to his instructional value system. No one else is in a position to make this decision for him.

No matter how effective an instructional system is, its ultimate impact on education will be determined by how generally, and how appropriately, it is used. Ideally, both instructional tools and instructional problems would be described in such a manner that they could be matched one with another appropriately. Unfortunately, this is not presently the case. Tools of instruction, whether books, filmstrips, or motion pictures, are usually described and evaluated in terms of their physical properties. Catalogues list the number of frames in a filmstrip, the length of a motion picture and whether it is black and white or color, and so for[illegible]. Yet the ultimate purpose of these tools is that they work some [illegible]ange in the students who use them. It would seem that a far more appropriate criterion for selection and a far more useful means of description would be the behavioral rather than the physical properties of these tools. The prudent car buyer will assess the performance characteristics of a car by taking it out on a trial drive. He will observe how it handles, the smoothness of the ride, and the quietness of the motor. The selection of instructional tools merits no less caution.

The ABCD's of Writing Behavioral Objectives

We have discussed the function of behavioral objectives at length to underscore the crucial role that they can play in the development of an instructional system. Now it is time to consider the form that these statements should take. As far as putting objectives into their appropriate form is concerned, the task can be described as simply writing sentences wherein each of the elements, subject, verb, and modifiers, has certain desired characteristics. We shall again use a mnemonic device, as the above heading indicates, to ensure that the information presented here, if not memorable, will be at least recallable.

Audience. The "subject" of the sentence used to describe a behavioral objective should describe who is to be doing the learning. The description should be in terms most relevant to the instructional task at hand. Sometimes this description will be simply an identification of the grade level and the subject of a class, for example, seventh grade geography students.

It is necessary to exercise care here, however, because the developer of the system is already making some commitments. Does the system really propose to achieve the behavioral objectives with all seventh grade geography students? Experience indicates that there are indeed few objectives that will be achieved by all students. For one reason or another, some students inevitably fall by the wayside. If the objective is to be attained by all the students, then the system will have to provide for students with low reading ability, low intelligence, or poor background in the subject matter. If the system is designed for students with average or better reading ability or intelligence, and certain required subject matter skills, then the "audience" should be so defined in the subject of the sentence.

While it is important to provide sufficient relevant information in describing the audience, it is almost as important to avoid being compulsive about it. Unnecessary details should be avoided. Learner populations can be described in great detail, if you put your mind to it, but it takes time to write such descriptions, and time to read them. The writer should only provide such descriptions as will be useful.

Behavior. It may seem self-evident at this point that the verb in a sentence that describes a behavioral objective is perhaps the most crucial element. A short cut used by those who edit and evaluate behavioral objectives is to skim through them, looking at the verbs. The use of inappropriate verbs is the most frequent, and most disabling, error in writing behavioral objectives.

The verb should describe an observable action that the studentswill demonstrate as a result of their learning experience. Some verbs denote readily observable actions, others do not. It is difficult to visualize just what a student looks like when he is understanding, appreciating, or even listening. But it is fairly easy to observe whether he is or is not speaking, writing, or constructing.

The following list of behaviors may prove to be helpful. Desmond Wedberg has asserted that all behavioral outcomes can be classified in one of these ten categories. Whether or not this is the case, the list is at least suggestive of useful verbs.

1. Identify
2. Name
3. Order
4. Demonstrate
5. Describe
6. Construct
7. State a rule
8. Apply a rule
9. Interpret
10. Distinguish

Conditions. It was indicated earlier that the purpose of behavioral objectives was not so much to focus our interest on what is observable, as to render more observable what is of interest. The degree to which we succeed in this will be determined largely by the care and ingenuity employed in describing the conditions of performance. It is here that we describe the setting for the evaluation, the materials and aids the student will be given, and the nature of the problems with which he will be confronted. It is important to note that this section of the objective does not deal with learning conditions, but with evaluation conditions.

If the behavior called for in the objective is to identify each state of the United States, then an appropriate phrase describing the conditions might be, "Given a blank outline map of the 48 contiguous states, on which only state boundaries are indicated...." But sometimes considerably more ingenuity will be required in devising evaluation situations. If the instructional objective is something as subtle as increasing interest in reading poetry, it may be necessary to describe a free reading situation, where the teacher systematically but unobtrusively notes how many of the students in the class choose to read poetry. Other conditions may be appropriate for the same objective. A situation could be described where poetry books along with others were made available to the students and a log was kept of the books checked out.

While the most frequent source of failure of an objective is inadequate description of behavior, the difference between an ordinary and an excellent objective is frequently the amount of ingenuity exercised in describing the conditions of performance.

Design. It seldom happens that the behaviors described in an objective are completely mastered by all students. To be sure, there are occasions when perfect mastery is required, and instruction will be continued until it is attained. But most often we are willing to tolerate a certain variability of performance and are satisfied with some reasonably high, but less than perfect, degree of attainment. This portion of the statement of objectives establishes the decision point where the instruction is deemed to have fulfilled its purpose, and a new objective can be pursued.

If the degree of attainment is set too high, it may require an extravagant use of instructional resources. It may even render the development of such an instructional system economically unfeasible. This is a matter that can be settled only by the careful weighing of a number of values. The experience of the instructor as to what constitutes a realistic expectation of students, and the extent to which mastery of this objective is prerequisite to subsequent attainment of other objectives, must be considered. When the developer of an instructional system prepares this portion of his objective, he should recall that it is he, not his students, who is committed to achieve it. Failure cannot be resolved by assigning grades. It requires a reworking of the instructional system.

Source Materials for Behavioral Objectives

The behavioral objectives for an instructional system do not emerge by a process of spontaneous generation, nor are they entirely the product of creative imagination. Fortunately, there are a number of sources and procedures that the writer of behavioral objectives can employ, that will simplify his task and contribute considerably to the outcome. A few of these will be mentioned here. The resourceful writer of objectives will probably discover some others.

It is usually advisable to gather information from several different sources, rather than depending too heavily upon one. For example, it will frequently be observed that a professor will verbalize one set of objectives, teach to a second, and test a third. One can only guess which set of objectives is the most valid. Choosing only one of these sources may well result in an unacceptable set of objectives.

The most useful sources of information will depend upon whether the instructional system is intended to replace some conventional instruction already being given, or is intended to develop new understandings and skills not presently taught; whether the objectives are primarily intellectual or manipulative; and whether the writer is working on his own instructional problem, or serving as a specialist for an instructional problem where the subject matter is outside of his area of competence. Following are some suggested procedures that should be helpful in most situations.

Verbalized Objectives. Perhaps the first stop in preparing a set of behavioral objectives is to interview the person or persons who understand best what the outcomes of the instructional system should be, or who have the greatest stake in its success. Even if one is writing objectives for an instructional system that he will build for himself, it is advisable to begin by expressing the objectives as articulately as possible.

A frequent mistake is to "behavioralize" the objectives too soon. This is particularly true when the writer is interviewing a subject matter specialist who is not accustomed to thinking in behavioral terms. Forcing the interviewee to speak in your terms, rather than trying to understand him on his terms, can easily lead to frustration and loss of cooperation.

An effective technique seems to be to gradually "zero in" on behavioral descriptions. Only after becoming familiar with the instructor's value orientations, his philosophical goals, and instructional aims, should one try to become more specific. The early part of the interview can supply many valuable clues as to what behavioral objectives will be considered most relevant. If the instructor is forced to supply behavioral descriptions too soon, he may merely narrow his attention to those outcomes most easily described, and never reveal those outcomes that are of most interest to him.

Once you have determined the "direction" in which the objectives lie, it is time to move in for a more explicit description. One can ask, "What does success look like?" "What does a student do that indicates to you that he has really learned what you wanted him to learn?" "If the student could do this and nothing more, would you be satisfied?" "What else could he do that would be just as good?" "How reliably should a student be able to do this? Say, how many times out of ten?" "Should every student be able to do this?" "What proportion of the students?"

The interviewer should have little difficulty generating meaningful questions of this sort, that don't confuse the interviewee with unfamiliar terminology. The answers will frequently come easily, but sometimes only with great difficulty. It will be helpful to explain why the answers are important.

It is usually neither feasible nor advisable to attempt to arrive at final statements of behavioral objectives in the interview, certainly not in the first interview. These can be written independently, and later offered for approval. Lengthy discussion about specific wording problems may arouse defensiveness, lead to futile digressions, and obscure more important information.

One should not expect too much from the interview, or the interviewee. Productive interviews are taxing at best, and much of the information required can and should be obtained from other sources.

<u>Instructional Materials and Activities</u>. Course syllabi and assigned reading materials can provide a considerable amount of raw material for writing objectives. They specify the elements to be learned, but usually reveal little about the way the student is intended to behave when he has learned them. They do, however, give the writer a comprehensive view of the range of objectives that will be required for the system.

It is often possible, by observing the activity of an instructor, to determine his appraisal of the relative value of different elements of subject matter, and to determine what kind of learning behavior he expects from his students. He may, for example, reveal little concern for detailed factual data, but emphasize analytical thinking. But the most explicit, if not the most reliable, indicator of the learning outcomes that the instructor considers important is the procedure he uses for evaluation.

Tests and Evaluations. Since test performances do represent behavioral outcomes, it would seem that they are a natural place to look for behavioral objectives. There is a certain amount of risk involved in this, because the nature of most tests is influenced heavily by factors other than instructional objectives. They often measure the most testable rather than the most significant outcomes. Few professors are trained in test construction, and few enjoy doing it. Quite naturally, considerably more energy is devoted to the preparation of instruction than to the preparation of tests. It is important to the instructor that a test be easy to construct, administer, and score, and that it yield a distribution of scores suitable for the assignment of grades. While such a test may certainly measure some of the desired learning outcomes, it is quite unlikely that it measures all of them.

It is also helpful to examine student products that have been evaluated. Not only do these yield positive and negative instances of the desired behaviors, but they facilitate the establishment of minimum standards of acceptable performance. The conditions of evaluation that are observed may suggest conditions that should be included in the statements of behavioral objectives.

Task Analysis. Particularly when the instructional system to be developed is directed at achieving psycho-motor skills, it is desirable to observe a number of individuals performing the task, both successfully and unsuccessfully. It is hard to analyze a complex performance into behavioral components if only successful performances are observed. But if one can observe a variety of unsuccessful performances as well, the various reasons for their

being deemed unsatisfactory suggest the variety of behavioral components required for successful performance. By thus breaking down one complex behavioral objective into a number of simpler behavioral objectives, the function of the objectives in facilitating planning and development of an instructional system is enhanced considerably.

Preparing the Statement of Objectives.

The procedures described above should yield more than enough information for writing the objectives. The task now becomes one of organizing and simplifying. Discrepancies and inconsistencies will have to be resolved, and behavioral statements examined to see if anything has been lost in the translation. Irrelevant data must be eliminated, and assumptions questioned. Even when this has been accomplished, the task of actually listing the objectives may seem formidable, particularly for more extensive instructional systems.

While the statement of instructional objectives must be both comprehensive and explicit, it does not follow that there must be a separate statement for each objective, nor is it necessary to repeat compulsively elements that a number of objectives have in common. For example, if all of the objectives are appropriate for a given population of students, the audience need be described only once in a prefatory statement. Objectives to be measured under similar conditions can be grouped together, and the conditions described but once. In some instances, it will not be feasible to measure all intended behaviors for all students. For example, if one hundred definitions are to be learned, a sample of these definitions can be tested and the proportion of that sample that constitutes an acceptable degree of achievement can be specified.

Introduction

"We learn to do neither by thinking nor by doing, but by thinking about what we are doing."

George D. Stoddard

This paper has urged that goals have primacy over procedures, and that the latter be evaluated in terms of their contribution to the former. It would indeed be an irony if the procedures suggested here would themselves be accepted uncritically and employed unquestioningly.

Effectiveness in writing behavioral objectives will be achieved from examined experience. What has gone before is merely a preface, meaningful only if it leads to such experience. Thus it seems more appropriate at this point to introduce, rather than conclude or summarize.

The procedures and rationale presented here are worthy of testing, but not blind acceptance. They are no more inviolable than the practice of beginning a paper with the Introduction.

RECOMMENDED REFERENCES

Ammerman, Harry L., et al. The Derivation, Analysis and Classification of Instructional Objectives. Washington, D.C.: Clearinghouse for Federal Scientific and Technical Information, 1966.

Ammerman, Harry L. Development of Procedures for Deriving Training Objectives for Junior Officer Jobs. Alexandria, Virginia: The George Washington University Human Resources Research Office, 1966.

Bloom, Benjamin S., ed. Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. New York: Longmans, Green and Company, 1956.

Krathwohl, David R., et al. Taxonomy of Educational Objectives, Handbook II: Affective Domain. New York: David McKay Company, Inc., 1964.

Mager, Robert F. Preparing Objectives for Programmed Instruction. San Francisco: Fearon Publishers, 1962.

Paulson, Casper F., Jr. "The Place of Prediction in Education," a paper presented in February, 1967, at the annual meetings of the American Educational Research Association as part of a symposium on The Prediction of Teaching Behavior.

III - OBJECTIVE ANALYSIS

Evaluation Sheet

We need to know how well the ideas and issues in this manual are communicated to you. You are the test audience for this material. To remove or strengthen the weak spots, to retain or improve the strong ones (if any), an account of your learning experience <u>as you read</u> these sections is crucial. Use this sheet as you study and jot down your reactions. Don't be concerned with typographical errors. We're concerned with how the message is coming through.

SECTION # ________ TITLE ______________________________

Give topic, paragraph or sentence, page	Importance Rate 1 2 3 4 5 Low - High	Clarity Rate 1 2 3 4 5 Low - High	Suggested Improvements

Over-all Rating of the Section: Use a 5-point Scale

Important				
1	2	3	4	5
Very Low				Very High

(Circle one)

Understandable				
1	2	3	4	5
Very Low				Very High

(Circle one)

PLEASE PUT ANY ADDITIONAL COMMENTS ON BACK

SECTION III
Objective Analysis and Instructional Specification

Table of Contents

SECTION III

Objective Analysis and Instructional Specification

Paul A. Twelker

Overview

In a systems approach to instruction, the systems designer (or "instructional engineer") has the task of creating and sequencing a series of learning experiences which will produce a previously stated behavior. To achieve this end, the first thing the systems designer must do is to define the instructional objectives or desired terminal performance outcomes. You are by now familiar with the processes involved in stating behavioral objectives. When a systems designer states what he wishes the student to learn in behavioral terms, he not only provides a means for looking at the instructional content in a facilitative manner, but he also provides a means for determining the criterion measures required to test whether or not the students have learned. By specifying the instructional content in behavioral terms, the designer provides himself with information that pertains to the design of instructional methods and materials. This information is obtained through processes which we shall call "objective analysis" and "instructional specification."

The instructional methods and materials, when implemented according to some administrative instructional plan, become the instructional program--after exhaustive try-out and revision. At the completion of instruction, criterion measures are applied to obtain an indication of the adequacy of the program. An integral step in the development of a new instructional system is the provision of self-corrective feedback all along the way that is used to modify the system if the desired effects have not been achieved. The learning experience is continually monitored and empirically developed until effective.

In summary, the process involved in designing an instructional system may be analyzed into four steps: (1) determining terminal objectives and sub-objectives (or enabling objectives) through objective analysis (What shall I teach?); (2) determining the instructional conditions (How shall I teach it?); (3) developing the instructional conditions (What's wrong with my prototype, and how can I correct it?); and (4) measuring the results of instruction (Did I teach it?). The topic of objective analysis falls within the first step, while the topic of instructional specification falls within the second step.

Predesign vs Extemporaneous Design

This section is designed to make clear the philosophical issues involved in the design of instructional materials in the manner described above. Frankly, we want you to be aware of these issues and at least accept our point of view long enough to understand it. In some future time, you may wish to try it out.

Instructional conditions may be predesigned or they may be designed "on the spot" while instruction proceeds. The latter method is employed by most teachers as "just telling or showing the student what you want him to know." "Extemporaneous design of the conditions of learning, during interaction with a student or students, is undoubtedly considered by many teachers one of their most important functions (7, pp. 251). Programed instructional materials, on the other hand, illustrate intentional predesign of the conditions of instruction.

There are arguments both for and against the predesign of instructional conditions. An obvious argument against predesign is that it is too time consuming for the average instructor. Loaded down with a full schedule of courses, an instructor simply does not have the time to develop an instructional system. An argument in favor of predesign is that the pre-selection of proper instructional conditions may be made in an unhurried manner rather than "on the spot." Think about the matter of predesign vs extemporaneous design. What are the advantages of predesign over extemporaneous design? What are the disadvantages? Consider the feasibility of each method. Jot your ideas down below.

Disadvantages of Predesign of Instructional Conditions

Advantages of Predesign of Instructional Conditions

Other Factors to be Considered

Gagné cites several important advantages of predesign of instructional conditions. They include the following:

"1. The selection of proper learning conditions may be made as an unhurried choice, rather than in 'spur of the moment' decisions.

2. A 'quality control' of the choice of learning conditions is ensured and maintained. Quality does not suffer from variations in teachers' skills.

3. Predesign makes possible pretesting. Whether or not a set of learning conditions has been correctly chosen and designed can be determined by trying it out on students, and revising if necessary.

4. Predesign of learning conditions greatly reduces the necessity for the teacher to use valuable time in extemporaneous design, and thus makes it possible for a proper emphasis to be restored to the teacher functions of <u>managing instruction</u>, <u>motivating</u>, <u>generalizing</u>, <u>and assessing</u>" (7, p. 253).

Try to find the time to read Gagné's discussion of predesign (7, pp. 250-254). It will give you greater insight into the advantages of predesign. Gagné claims that it appears to be completely feasible and well within the limits of an instructors' time restrictions. This is a point that sometimes provides a catalyst for fierce arguments in favor of extemporaneous design. It should be pointed out, however, that knowing <u>what to teach</u> (knowing the objectives of instruction and all the related subobjectives) brings the instructor a long way toward specifying instructional conditions and eventually provides the opportunity for empirical testing and "validating" the effectiveness of the instructional conditions.

In your list of disadvantages of predesign, you may have mentioned that extemporaneous design better takes into account the individual differences of students. You might have argued that with extemporaneous design, the teacher can detect differences of individual learning and react to them extemporaneously during instruction. Gagné contends that this is an unconvincing argument. He doubts that the typical classroom provides a place where the task of designing instructional conditions for individual students can even be accomplished. Because of high load requirements, the teacher simply can not deal with each student individually.

Further, extemporaneous design might produce instructional conditions that are too slow for the better learner, might move at the proper pace for the average learner, and it might miss the slower learner completely. Gagné also points out that predesigned instruction may take individual differences into account. You may wish to pursue some of these points later in the small group discussion period.

Relation Between Research and the "Instructional Systems Approach"

It should be noted that our present state of knowledge exhibits more of the characteristics of an art than a technology. However, we can not go so far as to say that there are no principles or procedures that might guide our efforts. Unfortunately, you might find that some of these principles may be more readily applied to specifying the instructional conditions for science courses than for humanities courses. This is simply to say that we cannot be dogmatic when it comes to developing instructional conditions. The procedures noted below are valuable only as they help the instructor reach his goal.

The guidelines presented below do not give the instructor hard and fast rules for specifying the series of learning experiences which will produce a previously stated behavior. What the guidelines are designed for is to help the instructor look at the problem of instructional specification in a systematic and, hopefully, beneficial manner. All of these guidelines may not be new to you. Certainly, a competent instructor utilizes many of the procedures listed below as he attempts to prescribe the learning experiences for a course. Further, these guidelines do not at all inhibit an instructor's creativity in specifying instruction. Rather, it is hoped that the guidelines will enable an instructor to be optimally creative. In other words, what is presented below are not answers for the instructor who is grappling with the problem of instructional specification--rather, they are suggestive of appropriate questions an instructor should ask as he sits down to determine what and how he should teach.

What relation does the specification of instructional conditions have to your development of research hypotheses? How can the procedures mentioned below concerning instructional specification help you in deciding what to research in terms of your own instruction. To paraphrase Gustafson (10, p. 10), neither objective analysis, instructional specification, or even the determination of training objectives can significantly advance the state of the art in education. Genuine improvements can

evolve only through empirical research conducted in accordance with sound experimental practices. However, the instructional systems approach can help clarify problems and bring the experience and ingenuity of instructional specialists to bear more directly on them. The proce[illegible] redesigning instruction, which must be repeated anew for [illegible] and each research project, has the dual effect of (1) [illegible] the application of subjectivity to only those areas where the state of the art makes subjectivity necessary, and (2) clarifying the issue so that the state of the art can be brought into play most effectively.

With the aid of an instructional specialist and a media specialist, the instructor brings to bear on the question of instructional specifications all that he knows, and all that is known by others (within personal limitations) about the subject. Only then may he conduct research in a meaningful manner. The manner in which research hypotheses may be derived is taught to every graduate student. Confronted with a problem, he lists all of the alternative hypotheses he can think of, and then tries to systematically disprove each hypothesis, either by citing theory, empirical evidence, or common sense. Those alternative hypotheses that are not discarded in this manner are examined more closely as probable hypotheses to research. This procedure, in effect, provides a powerful tool for the instructor who is grappling with the problem of instructional specification. With the aid of other experts, he attempts to discard as best he can all of the alternative hypotheses he has specified concerning an instructional problem he faces in his courses. The procedures of stating objectives in behavioral terms, deriving the sub-objectives, and determining the instructional conditions, aid the instructor in sensitizing himself to those areas of instruction which require his greatest attention.

With this procedure in mind, examine the following guidelines, and note how the procedure of setting up alternative hypotheses and systematically attempting to discard them fits within each guideline. The following section deals in the prerequisites that must be considered by an instructor before he designs an instructional system.

Prerequisites to Objective Analysis and Instructional Specification

A prerequisite to objective analysis and instructional specification involves the identification of "constraints" on the instructional system (these are prescribed limitations on what and how you can teach). For example, some obvious constraints are:

(1) course descriptions, (2) number of students, and (3) available finances. Take a few moments to think about other constraints that you as an instructor must consider as you specify instructional conditions for one of your courses. List as many of these constraints as you can.

Constraints

Answer:

> Some constraints that you listed might fall under these headings: (1) information sources, (2) resource personnel available, (3) equipment and materials available, (4) time limits, (5) characteristics of students, (6) facilities, and (7) trade-offs such as cost-effectiveness, time-effectiveness and so forth.

After an instructor has a clear idea of the nature of his students and the nature of the constraints placed on the instructional system, he is ready to determine what he is going to teach. Thus is the topic of the next unit.

Objective Analysis

The statement of instructional objectives is but a first step in determining instructional requirements and has been discussed elsewhere. Not only would it be desirable to check the initial statement of instructional objectives, it is also important to determine what the student specifically needs to learn. It is unrealistic to assume that the first attempt at the statement of behavioral objectives will result in a list that will be kept unchanged for any length of time. Something as important as objectives cannot be treated superficially. They are not written so we can file them away unchanged once they have been specified. In relation to the second point—determining what the student should learn—there must be some means to bridge the gap between where the student is at the beginning of instruction and where he should be upon completion of instruction. That is, we must specify "enabling objectives"—objectives that consist of the component actions, knowledges, skills, and so forth that enable the student to attain the terminal objectives specified. Enabling objectives may consist of basic factual and conceptual knowledge serving as background information necessary to attain the terminal objectives. The tool that we may use to specify the enabling objectives and the necessary background information may be termed "objective analysis." This analysis of subordinate behavior is based upon procedures used by Gagne which is sometimes termed hierarchical analysis or learning set analysis (6, 7, 9).

For some subject matters, the enabling objectives or subordinate knowledge may be arranged in a pyramid type of structure

as shown in Figure 1. In such a structure, layers of competencies are identified. The basic procedure is simple. Starting with the terminal objectives stated in behavioral terms, the following question is asked: "What kind of capability would an individual have to possess if he were able to perform this objective successfully, were we to give him only instructions?" The phrase regarding instructions needs some clarification. The learner should be: (1) told the form of the answer (e.g. numerical, symbolic, or enacted); (2) informed of any definitions that would clarify the meaning of the objective; (3) provided with guidance suggesting the application of previously acquired information to the objective under consideration.

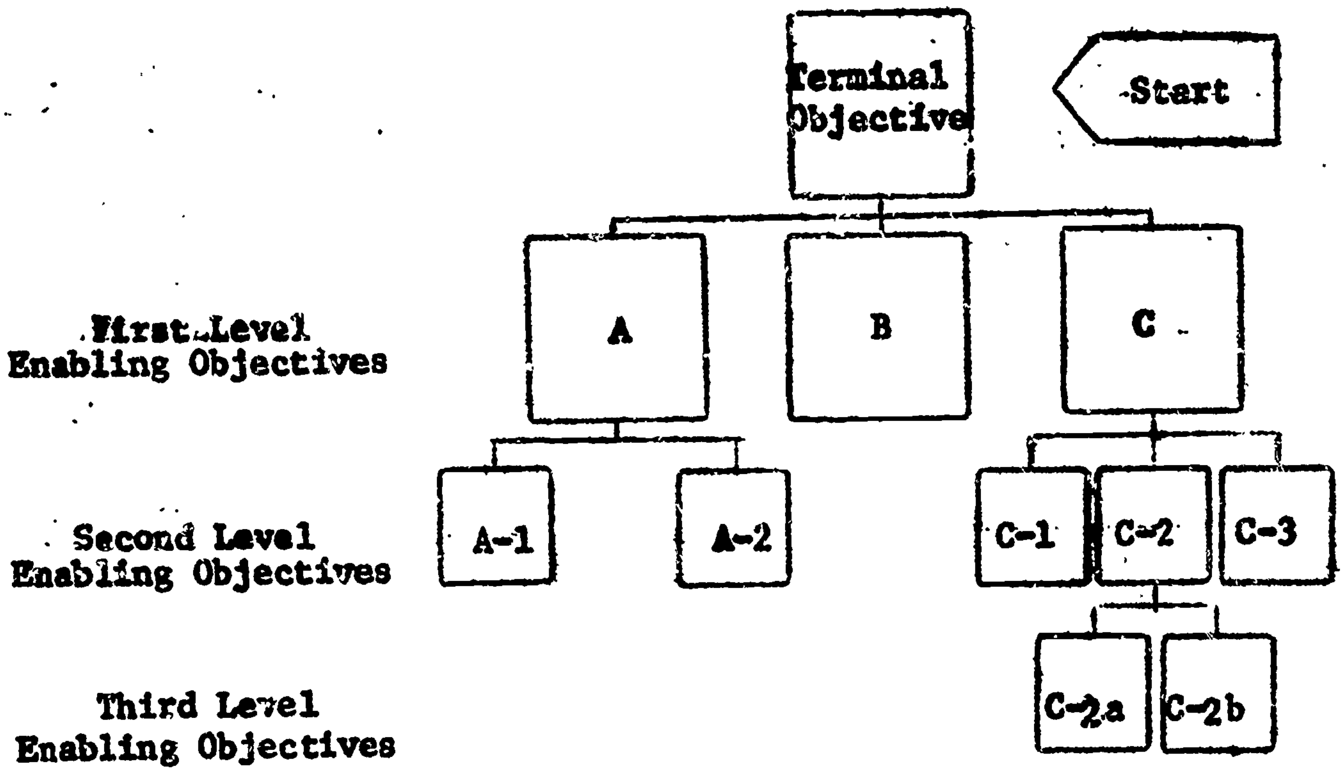


Figure 1. A hypothetical hierarchy of enabling objectives for a single terminal objective.

In the case of the hypothetical example in Figure 1, the answer to the question, "What would the individual have to know?", it turns out, identifies a small list of major items of knowledge or skill. These are represented by A, B, and C, on the figure. Then the same question is asked for each of these items in turn, and so on until each item of knowledge and skill derived from the objective analysis has been added to the structure. Note that the enabling objectives thus gained are not the same performance as the final task from which they were derived. They are in some sense simpler and in some sense more general. By using the

procedure of objective analysis, we find out that what we are defining is a hierarchy or suborder of knowledges or skills, growing increasingly simple. In such a structure, options exist within layers as to sequence of acquisition, but acquisition of one layer in its entirity is called for before the presentation of the next higher or more complex layer. For example, A-1 might be taught before or after A-2 is taught, but each of these must be taught before A is taught. Some courses involve instruction where the structure of content is relatively flat, while other courses represent hierarchies of skills as shown here. However, in both hierarchical and flat structure there may be many options possible in the sequencing of the objectives for the course.

An example of a hierarchy of enabling objectives for a topic in elementary non-metric geometry is shown in Figure 2. The topic consists of knowledge which a student can use to specify sets, intersections of sets, and separations of sets. Note that at the lowest level (VI and V), the concepts of separation of entities into groups, point and set of points, must be known or acquired before the learner is ready to attain the higher order enabling objectives. Take a few minutes to study the hierarchy and prove to yourself that the lower order objectives must be attained prior to the higher order objectives.

You may also wish to examine the question of what happens when learners actually undertake to acquire a set of principles that appear to have this hierarchical structure. Does the mastery of a lower order objective actually affect the learning of the next "higher" objective, as would be expected? This question was tested in a study with sixth graders. The results of the study are summarized below.

> "The results showed that the learning of 'higher-level' principles was dependent on the mastery of prerequisite 'lower-level' principles in a highly predictable fashion. For example, of the 72 students who performed correctly on Principle IIa, only one did not perform Principle IIIa correctly on the test. Of the 18 students who did Principle IIa incorrectly, all 18 did Principle IIIa incorrectly. The prediction that learning IIa depends on knowing IIIa was borne out, therefore, with a frequency of 99 per cent. For all the other possible comparisons, . . . the frequency of correspondence between predictions and findings ranged from 95 to 100 per cent. The learning of organized knowledge, according to these results, appears to be predictable from the pattern of prerequisite principles that make up the hierarchy of knowledge to be acquired" (7, p. 153).

The conclusion cited above has been verified in a number of studies using topics in mathematics (6, 8, 9).

Before going on to another example, take another look at the hierarchy in Figure 2. Find the box labeled IIc, "Identify and draw the intersections of lines or parts of lines taken two at a time, as more than one point." Below, list the enabling objectives that the student must know how to do in order to be instructed in this principle. Include only those enabling objectives that are pertinent for the student attaining IIc.

<u>IIc</u>

<u>Enabling Objectives</u>

____

____

____

____

____

____

____

____

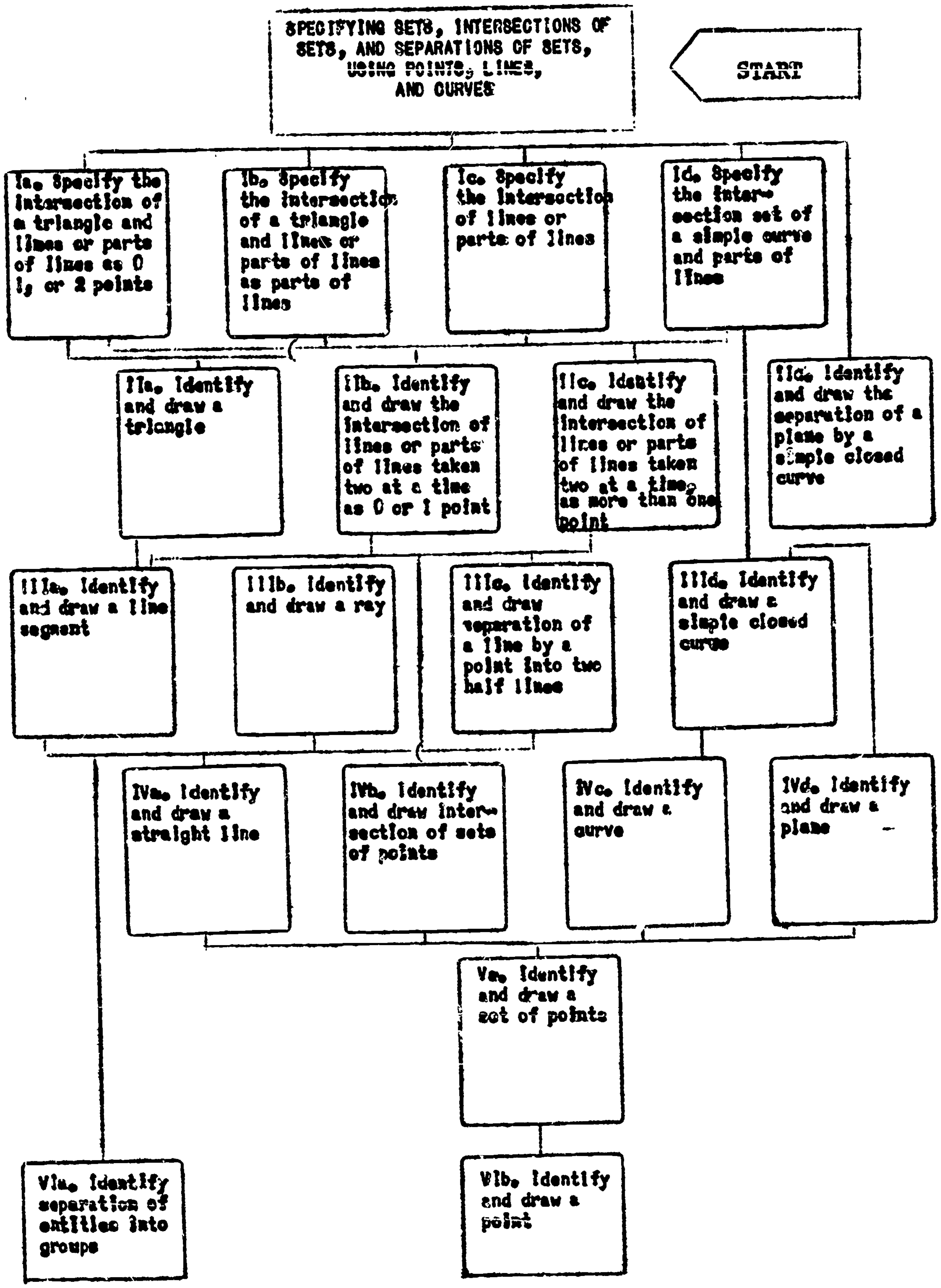


Figure 2. (Caption on page 13).

Answer

> By asking the question, "What would the student have to know how to do in order to be instructed in this principle?", you will find that he must know principles IIIa, IIIb, and IIIc. Further he must know IVa, IVb, Va, VIa and VIb because IIc, the higher-order principle, incorporates these lower-order principles in the sense that the learner must be able to perform these operations before he can progress to the next level.

Another hierarchy of knowledge for two tasks is shown in Figure 3. Take a few minutes to trace through the hierarchy. Identify all of the lower-order enabling objectives necessary for the student to know in order to satisfy IIa and IIb. List these objectives below.

IIa	IIb
Enabling Objectives	Enabling Objectives
________________	________________
________________	________________
	________________
	________________
	________________
	________________

Figure 2. Hierarchy of enabling objectives for a topic in elementary non-metric geometry. Adapted from Gagné (7, p. 150).

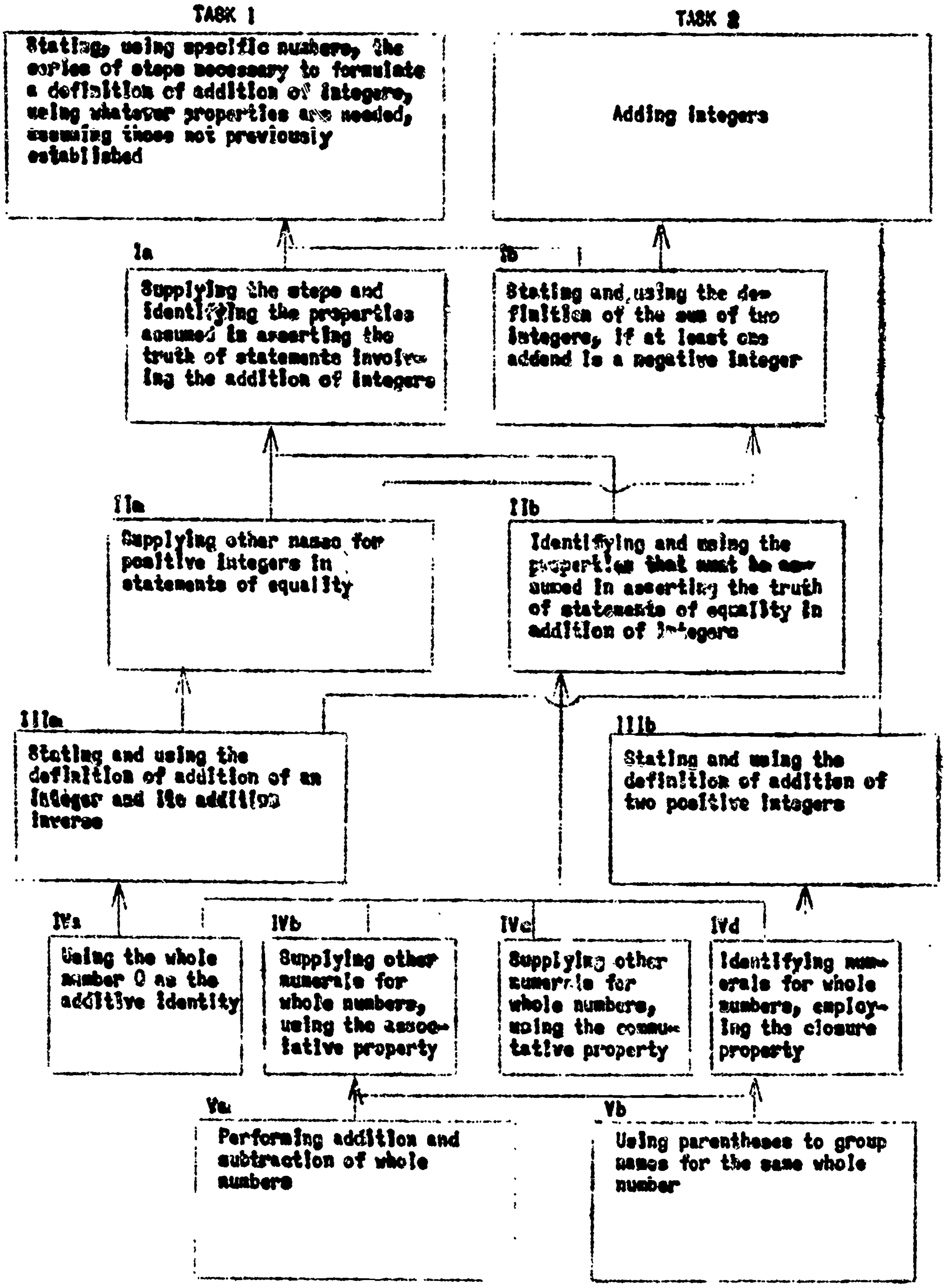


Figure 3. Hierarchy of enabling objectives for a topic in mathematics. Adapted from Gagné, et al. (8)

Answer

> In order to accomplish IIa, the learner must be able to state and use the definition of addition of an integer and its additive inverse (IIIa) as well as using the whole number 0 as the additive identity (IVa). In order to accomplish IIb, the learner must be able to accomplish all those behaviors stated in Levels IV and V.

Now take a look at Figure 4. Examine the terminal objective and note that three components must be known before the student can perform the objective. The first (Ia) requires that the student be able to identify the forces acting on a body in opposition to each other. In order to learn this principle, the student must first know IIa, that is, the conditions for equilibrium.

Examine the two enabling objectives listed at the bottom of the figure. Then ask the question, "What would the student have to know?" for objective Ic. Which objective fits in the frame labeled IIb?

____

Which objective is a prerequisite for IIIa?

____

Check your answers by examining principle Ib. Does the enabling objective you picked as IIIa represent a prerequisite principle for Ib and not for Ic?

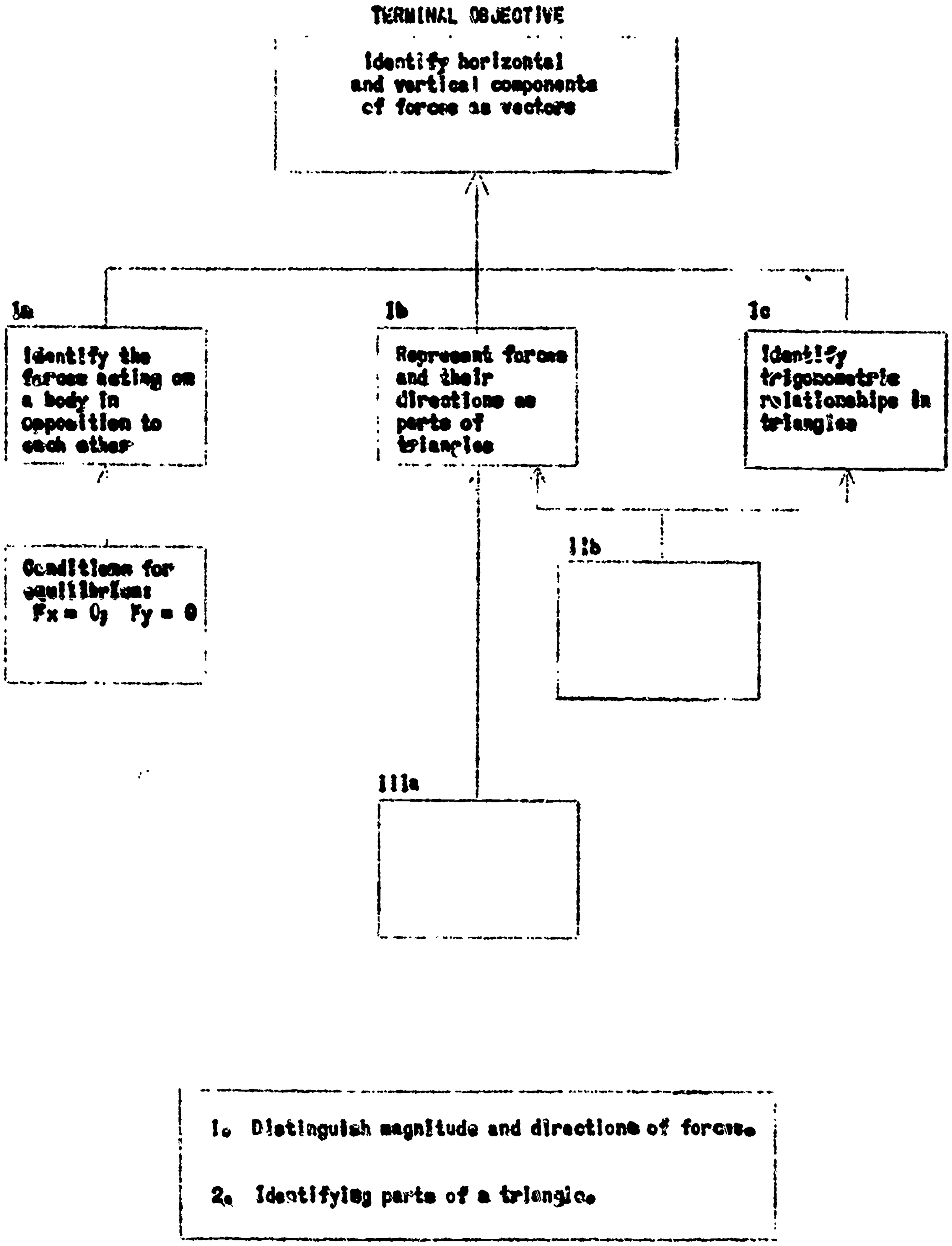


Figure 4. Hierarchy of enabling objectives for a physics topic. (Adopted from Gagné, (7, pp. 155).

Answer:

> Objective 1 belongs in the frame labeled IIIa and Objective 2 belongs in IIb. Being able to identify parts of a triangle is certainly a prerequisite to identifying trigonometric relationships in triangles.

Below are listed five enabling objectives. If they look familiar, it is because they are taken from Figure 2. But don't look back at that figure until you have tried ordering these objectives. Try to consciously ask the question, "What would the student have to know?" with each objective rather than simply ordering the objectives from the complex to the specific. Order these objectives beginning with the terminal objective.

	Order
Specify the intersection set of a simple circle and parts of lines	Terminal Objective (1)
Identify and draw a set of points	____________
Identify and draw a simple closed curve	____________
Identify and draw a point	____________
Identify and draw a curve	____________

The correct order appears in Figure 2. Check your answer by tracing the hierarchy for Objective Id.

It is a relatively simple task to order enabling objectives, asking the question, "What would the learner have to know?" It is a more difficult task to derive the enabling objectives from your own objective analysis. Rarely will you be fortunate enough to have the enabling objectives already formulated and diagramed for your particular terminal objective. Take a few minutes now to attempt an objective analysis on a behavioral objective of your own. These will be discussed during the small group discussion period. If you have any questions as a result of this exercise, jot them down. Bring them up during the small group discussion session.

Sometimes the question is asked, "If an hierarchical analysis is done correctly, can there be more than one structure obtained?"

To quote from Gagne, "There are perhaps several possible learning set hierarchies which could be worked out . . . , and it is quite conceivable that some are 'better' than others in a sense of being more efficient or more transferable to later learning" (9, p. 5). In the long run, any hierarchy that is developed is "good" only to the extent that it aids the instructor in determining what to teach, and in what sequence to teach it.

Another way to describe the organization of enabling objectives is by using an outline. The terminal objective diagrammed in Figure 1 could be outlined as follows:

I. Terminal Objective

  A. First level enabling objective
    1. Second level enabling objective
    2. Second level enabling objective

  B. First level enabling objective

  C. First level enabling objective
    1. Second level enabling objective
      a. Third level enabling objective
      b. Third level enabling objective
    2. Second level enabling objective
    3. Second level enabling objective

What are the drawbacks in representing all hierarchies in outline form? (Hint--see Figure 2 or 3).

Answer:

As shown by figure 2 and 3, some lower-order objectives "tie in" to more than one higher-order objectives. A hierarchy that is diagramed shows these relationships more easily.

It is most useful in organizing objectives to type or print each component on a 3 x 5 inch card and sort these cards on a large table. (You may find the chalkboard too inflexible to be of much value.) When the objectives have been arranged, they may be diagramed.

Why go through the rather tedious procedure of objective analysis? One reason for arranging enabling objectives in a sequence of some sort at the outset is to consider the extent to which learning of one objective may facilitate the learning of another, or at least may have a bearing upon the design of instruction to be specified. By laying out the structure of enabling objectives in such a manner, the instructor may readily determine if he has been teaching too little or too much. Teaching too little will cause the student to require further training while teaching too much might raise the cost of instruction and require longer course time. After an objective analysis is completed, it makes good sense to ask the question, "Can students perform the action in the terminal objective without this particular skill or knowledge?" If the answer is in the affirmative, that skill or knowledge may not be an enabling objective, at least not for some students. Enabling objectives are derived from a knowledge of both the terminal objectives and the existing capabilities of the student population. In some courses, the instructor cannot assume any abundance of prior knowledge at the start of instruction, while in other courses, (for example, graduate courses) the instructor assumes an abundance of prior knowledge. The limit to which an instructor details the structure of the objective analysis depends upon the entry behavior of his students.

A second reason for conducting an objective analysis is that it serves as a self-corrective means to check the behavioral objectives as initially stated. The instructor has a means of testing whether each objective is feasible to teach in a given amount of time and in the context of learning experiences designed for other objectives. For example, it is quite feasible that basic knowledge components are prerequisites for more than one objective. Knowing this, an instructor may capitalize on this fact and teach these particular objectives in sequence.

When an instructor or researcher has taken the time to conduct an objective analysis, he is well on his way toward knowing what to teach, and in what sequence to teach it. Objective analysis is a technique that, when mastered, may be applied by the instructor with little or no help from others, save an occasional conference with a colleague to check on questions concerning the subject matter. When an objective analysis is completed, the instructor essentially has determined what to teach, and is in a position to specify how to teach it. This is the topic of the next unit.

Determining The Instructional Conditions

Identifying for Each Objective the Type of Learning Represented. It would seem that an important basis for matching instructional objectives with instructional conditions would be to draw upon the large amount of information on human learning that has accumulated through the years. If each objective could be classified into a category which is homogenous with respect to the conditions fostering learning of that type, the instructor would have at his command a very powerful tool with which to specify instructional conditions. Eckstrand (5) has suggested that such a taxonomy would provide "a system for organizing existing information in a way which would facilitate its application to particular training problems" as well as provide "a most useful tool in determining deficiencies in our knowledge and thereby serve to guide future research." Several attempts have been made at specifying this learning taxonomy (e.g., 4, 7, 12.) Although each of these taxonomies leave much to be desired, Gagné's taxonomy seems to be the most promising.

Gagné (7) has identified eight types of learning. As these are identified below, keep in mind that the purpose of this taxonomy is to provide a system for specifying as best we can the different set of conditions necessary for the learning of each of the objectives previously identified.

The eight types of learning that Gagné distinquishes are as follows:

> Type 1: Signal Learning. A general, diffuse response is made to a signal (i.e., the classical conditioned response of Pavlov).
>
> Type 2: Stimulus-Response Learning. A precise response is made to a discriminated stimulus (i.e., trial-and-error learning, and instrumental learning).
>
> Type 3: Chaining. A chain of two or more stimulus-response connections are acquired by the learner (e.g., procedural skills).
>
> Type 4: Verbal Association (a subvariety of Chaining).

Verbal association involves the learning of chains that are verbal (e.g., paired associate learning such as the naming of objects).

Type 5: Multiple Discrimination. Different identifying responses are made to as many different stimuli, which may resemble each other in physical appearance to a greater or lesser degree. It is simply a matter of establishing numbers of different chains when the connections tend to interfere with each other's retention.

Type 6: Concept Learning. A capability of making a common response to a class of stimuli that may differ from each other widely in physical appearance is exhibited in concept learning.

Type 7: Principle Learning. Principles are chains of concepts that are generally termed knowledge. Principles represent the relationships among concepts.

Type 8: Problem Solving. The kind of learning that requires "thinking" may be termed problem solving. Essentially, two or more previously acquired principles are used to produce a new "higher-order" principle that results in a solution to a problem.

It would be impossible in the short time we have to discuss each of the eight types of learning in detail, and to expect you to master each type sufficiently so as to recognize examples or to generate your own examples. In fact, instructors faced with the task at hand will usually secure the services of an instructional specialist who is well acquainted with the various types of learning and the conditions necessary to bring the learning about. For purposes of this institute, however, it seems appropriate to at least look at the more important types so that you will have a better understanding of the processes involved in designing an instructional system.

In college instruction, the most frequently occuring types of learning probably are concept learning, principle learning, and problem solving. This is not to say that the simpler types of learning do occur in college instruction. For example, in foreign language classes, multiple discriminations may have to be established. Training certainly may occur in activities such as

pronouncing foreign words.

Below are descriptions of the three most frequently occurring types of learning. They have been taken from Gagné and simply present the highlights of the characteristics of each type of learning.

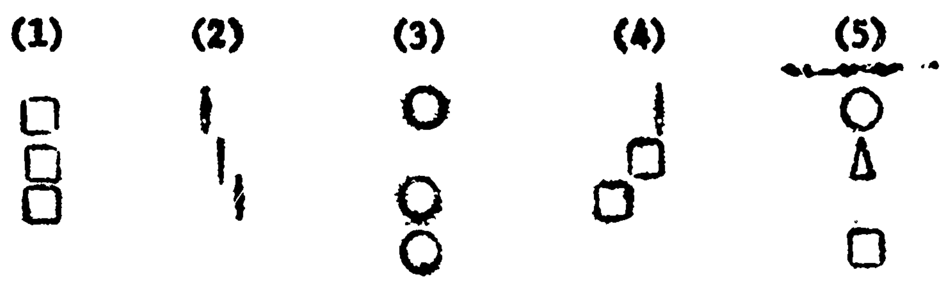
"Learning Type 6: Concept Learning

Attention now turns to a kind of learning that appears to be critically dependent on internal neural processes of <u>representation</u> for its very existence. In man, this function is served by language. Although a number of animals have been shown to possess the capacity to make internal representations of their environments, present evidence suggests that this capacity is extremely limited even in the higher apes. Human beings, in contrast, employ this capacity freely and prodigally; they are highly inclined to internalize their environment, to 'manipulate' it symbolically, to think about it in endless ways.

Learning a concept means learning to respond to stimuli in terms of abstracted properties like 'color,' 'shape,' 'position,' 'number,' as opposed to concrete physical properties like specific wave lengths or particular intensities. A child may learn to call a two-inch cube a 'block,' and also to apply this name to other objects that differ from it somewhat in size and shape. Later, he learns the concept <u>cube</u>, and by so doing is able to indentify a class of objects that differ from each other physically in infinite ways. A cube may be represented concretely by objects made of wood, glass, wire, or almost any material; the object may be of any color or texture, and of any size. Considering this great variety of physical stimulation that may correctly be identified as 'cubical,' perhaps it is not surprising that it takes some very precise language on the part of geometers to define what is meant by 'cube.' But, of course, a person does not have to understand such a definition in order to

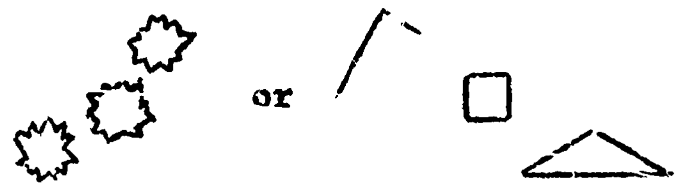
identify correctly a cube under most ordinary conditions of his existence. Except for some very special purposes of mathematical theory, an individual identifies a cube 'intuitively,' that is, on the basis of an internalized representation that does not employ the words of the geometer's definition. Whatever the process may be, there can be little doubt that a concept like cube is learned, and that its possession enables the individual to classify objects of widely differing physical appearance. His behavior comes to be controlled, not by particular stimuli that can be identified in specific physical terms, but by abstract properties of such stimuli.

As an example, we may consider how a child learns the concept middle. Initially, he may have been presented with a set of blocks arranged like this: □ ■ □. If previous Ss→R learning has enabled him to receive reinforcement for a request such as 'Give me a block,' he can then readily learn the simple chain of picking up the middle block when his parent says 'Give me the middle one.' Similar chains can then be established with other objects, such as balls arranged in the same configuration ○ ○ ○, or sticks | | |. One might want to make certain that the chains were generalizable over a range of separations, like | | | and | | |. In other words, the deliberate attempt is made to establish a number of chains applicable to a variety of specific physical configurations. Continuing this effort, the spoken word 'middle' might be applied to various other arrangements of objects, including such situations as these:



By this means, the child comes to respond correctly to middle as a concept meaning 'an object between two others.' (Other meanings of 'middle,' such as the 'center of an area,' are, of course, different, but may be similarly learned.)

How does one know whether the child has in fact learned the concept middle? The crucial test is whether he will be able to respond correctly, not by chance, to some new configurations of objects he has not previously used in the course of learning. For example, these might be:



If the child can now respond properly when told 'Give me the middle one,' it may be concluded that he has learned a concept, and that his behavior is not controlled by specific stimuli,' (7, p. 47-49).

"Learning Type 7: Principle Learning

As an approach is made to a description of the kind of learning that can be identified as most broadly applicable to the content of formal education, the job grows somewhat easier. The most difficult kinds of learning to describe are the simplest kinds, perhaps because it is necessary to be so careful to keep them simple. A kind of learning like principle learning has some well-known conditions for its establishment, which all of us employ very frequently. Continuation of the basic theme of the argument--that different varieties of lea ing can be distinguished by the conditions required to bring them about--leads to the view that principle learning is not only a highly familiar but also readily understandable kind of learning. As is true with certain forms, the most important conditions are the prerequisities for learning.

Principle learning may be exemplified by the acquisition of the 'idea' contained in such propositions as 'gases expand when heated'; 'the pronoun each takes a singualr verb'; 'salt is composed of the elements Na and Cl'; '$xa + xb = x(a + b)$'; 'the definite article die goes with a feminine (German) noun'; and many, many others. Surely there can be little doubt that human beings must learn large numbers of such principles, some interconnected, some not, along the road to attaining the status of being considered educated adults.

Possibly the immediate demurral will be raised, "Why, these are simply verbal facts to be memorized!" Not so; and that is why the word 'idea' has been employed to describe these principles, inexact as that word may be. From our previous discussion, it is apparent that each of these statements can be learned as a verbal chain. If we want to teach a five-year-old to memorize them in that way, this surely can be done. But the only kind of performance that would be possible following such learning would be something like this: Complete the statement: 'The pronoun each takes ___ _______ _____.' Such a performance is by no means what is established when one has learned a principle. In fact, one may learn the principle about 'each' and its verb without being able to verbalize such a statement at all. What is meant by learning such a principle is this capability: being able to use the singular form of the verb in a variety of sentences or clauses having each as a subject. If an individual is able to demonstrate this capability in a number of instances of representative sentences and clauses, one is justified in concluding that he has learned a principle.

In a formal sense, a principle is a chain of two or more concepts. Some would say, it is a relationship between concepts; but it seems preferable to state the nature of this relationship. The simplest type of principle may usually be cast in the form "If A, then B," as in the

example 'If a feminine noun, then the article __die__.'" (7, p. 51-52).

"Learning Type 8: Problem Solving

Once he has acquired some principles, the human being can use them for many purposes in dealing with and controlling his environment. He can also do something else that is most important: he can think. Basically, this means he is able to combine the principles he has already learned into a great variety of novel higher-order principles. He may do this by stimulating himself, in part, and also by responding to various forms of stimulation from his environment. By means of the process of combining old principles into new ones, he __solves problems__ that are new to him, and thus acquires still a greater store of new capabilities. The problems he solves are new to him, but they may not be new at all to other people. Of course, every so often a scientist or other creative person may arrive at a problem solution that is novel to society in general.

Problem solving, by which is meant 'thinking out' a new principle that combines previously learned principles, is a process that is very familiar to most adults. There is nothing very 'special' about such events, since they are likely to occur with a frequency of ten or twenty times daily in the life of an average man. When a driver maps his route through traffic (as opposed to simply being swept along by it) he is solving a problem. When a man replans his luncheon schedule as a result of a new appointment, he is solving a problem. When a housewife decides to shop selectively for certain items on the basis of differential prices, she is solving a problem. These everyday examples bear a close formal resemblance to the problems that are solved by students in composing reports and themes, in marshaling arguments to present a point of view, in performing laboratory experiments, and in reading a properly written textbook.

Suppose that a student of physics encounters this situation in reading his textbook:

> Power is defined as the work done divided by the time during which the work is done. That is,
>
> $$\text{Power} = \frac{\text{Work}}{t}$$
>
> Suppose we have a body doing work on another body, acting with force F and velocity v. Can you show that the power delivered to a body is the product of the force acting on it and its velocity?

The solution of such a problem requires thinking, which might go somewhat as follows. First, the student formulates to himself the fact that he wants an expression relating power to force (F) and velocity (v). He then recalls that work is related to force and the distance through which it moves, that is,

$$\text{Power} = \frac{\text{Work}}{t} = \frac{Fs}{t}$$

and he is therefore able to state that

$$\text{Power} = Fv$$

and to recognize this as the formulation he is seeking. Involved in these steps, of course, is the use of other principles he has previously learned, pertaining to the substitution of equivalent quantities in equations.

When these steps have been accomplished, it is reasonable to suppose that the student has learned something new. He has not learned simply that Power = Fv, because he could readily have learned that as a verbal sequence (type 4). Rather, he has learned how to demonstrate in concrete terms the relationship of power to force and velocity, beginning with the definition that power is work done per unit time. In other words, he has learned how to _use_ this

definition, and how to generalize it to some considerable degree to situations in which a force acts against a body with a known velocity. This act of problem solving, then, has resulted in some very substantial learning. The change in the individual's capability is just as clear and unambiguous as it is in any other variety of learning," (7, p. 54-56).

In a short paragraph, state the main distinctions between concept learning, principle learning, and problem solving.

Answer:

> In summary, we may say that when a learner responds to things or events as a class, what is involved is concept learning. Two or more concepts may be linked together to make up principles. Principles represent the relationships among concepts or among other principles. When the learner uses principles to achieve some goal, or when a higher-order principles is generated from other principles, problem solving is involved.

Below are listed several examples of each of the three types of learning discussed. Attempt to classify them as either: (1) concept learning, (2) principle learning, (3) problem solving. Several of the examples are shown with the particular type of learning represented to help you out.

1. A child learns to call square objects made of metal, glass, wood, and other material, "cubes."

   <u>concept</u> <u>learning</u>

2. A student uses the singular form of a verb in a variety of sentences or clauses adding <u>each</u> as a subject.

   <u>principle</u> <u>learning</u>

3. The student demonstrates that the power delivered to a body is the product of the force acting on it and its velocity.

   <u>problem</u> <u>solving</u>

4. A student can select a ruler and count the number of times it can be laid end to end along a linear dimension to be measured.

   ________ ________

5. A child points out objects that are blue from many objects of different colors.

   ________ ________

6. A child shows that he can use a ruler to determine whether two objects are the same length when he is not allowed to compare their lengths directly.

   ________ ________

7. A learner explains the fact that the bones of a whale have been found in the Sahara Desert.

_______ _______

8. A teacher points to an object that a student has not previously seen. The instructor asks, "Where is the edge?" and the student points to it. This is an example of

_______ _______

9. A student formulates an answer to the question, "Why does hot air rise?"

_______ _______

10. A student is presented with a number of objects, some of which are round and some of which are square. The instructor asks, "Which of these will roll?" and the student responds by rolling the round object.

_______ _______

11. Given instructions to do so, a student identifies and draws a point. His behavior shows that he knows the _______________ of "point."

12. A student answers the question, "In how many ways can identical jobs be filled by selections from 12 different people?"

_______ _______

13. A student lists each different ordering of a set of objects. He shows that he knows the _______________ of "permutation."

14. A student is asked to distinguish between two lists, one of which includes permutations and the other, combinations. He does so correctly, and has shown that he knows the _______________ of "permutation."

15. A student is shown and identifies several examples of striated muscle. He is then shown a new example and he responds, "This is striated muscle." He has shown that he knows the _______________ of "striated muscle."

16. A student is asked to demonstrate and describe how striated muscles work. Using a model of an arm and the muscle connections, he shows the movements made when the

muscle contracts. He shows that he knows the__________ of how striated muscles move an arm.

17. A learner is asked to mark the muscle connections and the direction of movement of limbs of a frog. This is an example of ________ ________.

Answer:

Check your answers with those listed below:

4.	Problem solving	11.	Concept
5.	Concept learning	12.	Problem solving
6.	Principle learning	13.	Principle
7.	Problem solving	14.	Concept
8.	Concept learning	15.	Concept
9.	Problem solving	16.	Principle
10.	Principle learning	17.	Problem solving

Identifying the Instructional Events that Provide General Conditions of Learning for Each Objective. After identifying the type of learning represented by each enabling objective, the task is to develop an instructional environment which will transform learners into graduates who can perform at the specified levels after instruction is terminated. To accomplish this, instructional sequences should be designed that reflect the conditions of learning previously identified for each objective.

A summary of the conditions for principle learning as cited by Gagné are listed below. If you wish to study the conditions of principle learning in greater detail, a discussion taken from Gagne is presented in Appendix A. For discussions of conditions of learning for the other types included in Gagné's taxonomy, refer to the book, The Conditions of Learning, (7).

"The requirements for instruction of principles whether practiced by a teacher, a film, or a text book (are):

Step 1: Inform the learner about the form of the performance to be expected when learning is completed.

Step 2: Question the learner in a way that requires the reinstatement (recall) of the previously learned concepts that make up the principle.

Step 3: Use verbal statements (cues) that will lead the learner to put the principle together, as a chain of concepts, in the proper order.

Step 4: By means of a question, ask the learner to "demonstrate" one of more concrete instances of the principle.

Step 5: (Optional, but useful for later instruction): By a suitable question, require the learner to make a verbal statement of the principle" (7, pp. 149).

As a college instructor, you probably do not teach principles in isolation as described above. Instead, you teach a number of principles, often without ever stopping for a breath in between. Consider, for example, the principle stated in Step 1 above. Rarely would an instructor question the learner to make sure that he knew the concepts of "inform," "learner," "form," "performance," and so forth. Indeed, the instructor usually does not even assume that the learner would not know these concepts. But this is precisely where misunderstandings occur in instruction. If a student does not understand even one concept included in a principle, the probability of his success in understanding that principle (demonstrating one or more concrete instances of the principle) is lower than if he knew that concept.

The examination of the conditions necessary to bring about learning of principles has been presented simply as an example of what an instructor, together with an instructional specialist, might go through in specifying instructional conditions. You are encouraged to study Gagne's book, The Conditions of Learning, for detailed discussions of the conditions of learning of each type discussed, as well as a book by Briggs (1) which gives concrete instances of the application of the techniques discussed.

Identifying the Form of the Instructional Event. At this point, decisions have to be made as to the nature of the instructional events required. For example, should they be verbal or non-verbal, and should they be visual or auditory. Of course, there are other considerations in special cases such as the requirement for motion, duration of exposure, and so forth.

Below, for each instructional event involved in principle learning, are listed some options as to the form of the instructional event. The example is taken from Briggs (1, pp. 33).

Instructional Event	
1. Inform learner about performance required.	May require objects or pictures (depending upon objective).

2. Stimulating recall of component concepts.	May be done by oral or printed speech; actual objects or pictures may be desirable (depending upon objective).
3. Verbal cueing.	Oral or printed words.
4. Appraisal.	May require objects or pictures (depending upon objective).

Let's examine an example given by Briggs (1, pp. 60-61) in detail. Although the example involves kindergarten children, the steps shown are applicable at all levels.

"Step 1

Objective: Demonstrate a method for using a measuring stick to determine whether two objects separated by space, are the same length.

Step 2

Type of Learning: Principle learning: 'Two objects equal to the same thing are equal to each other.'

Step 3

Conditions of Learning:

(a) Individual concepts must have previously been learned ('longer,' 'shorter,' 'same length').

(b) The principle must be stated to the child and demonstrated, or it must be discovered by the child.

(c) Several examples of application should be experienced to be sure that more than the verbalization of the principle is learned.

Media Program:

Instructional Event	Medium
1. Inform learner about performance required.	Two sets of dowels are mounted on separate wall displays. Each set has members of different lengths. Problem is how to identify the members of the same length in the two sets. (They cannot be moved.)
2. Stimulate recall of components concepts.	Child is led to recall how to find out if lengths are equal by side-by-side placement of objects.
3. Verbal Prompting	'Here are some loose dowels that can be moved. Would they help?'
4. By discovery or prompting is stated: If a movable dowel is equal to two others, as shown by side-by-side placement, the two others are the same length.	Children use loose dowels to make side-by-side comparisons to identify equal members in the two display sets of dowels.

For this objective, also, media alternatives may be selected for either individual or group instruction. For group instruction, the two displays and the loose dowels are viewed and matched. Verbal statements by the teacher present the performance sought, and either provide or prompt the statement of the principle. Children make several matchings, using different movable sets of dowels, straws, pencils. (Again, a motion picture might present this principle, but having real objects to manipulate might accelerate this application of side-by-side placements.)'

Recall that up to this point, the instructor is still working in cooperation with an instructional specialist. Once the

instructional program has been developed to this point, the instructor can proceed on his own to develop the program in terms of the actual words used to convey information. Of course, he will probably use the services of a production specialist if slides, overhead transparencies, or movies are required.

Here is another example showing how the instructional conditions are specified, given the conditions of learning. This example has been taken from Briggs (1, pp. 94-95).

Principle: The sun rises in the east because the earth rotates toward the east.

Media Program:

Instructional Event	Medium
1. Inform learners of performance required	This picture will enable you to explain sunrise, sunset, and time zones.
2. Present stimulus.	Motion picture of earth with cities labeled. Rotation of earth brings sunrise first to New York, then to Denver, then to San Francisco. Narration and arrows illustrate sunrise, noon, sunset, time zones.
3. Present prompt.	Movie sequences of the memory device illustrating "Rule of the right hand," which mediates recall of rotation toward the east.
4. Appraisal.	Programed instruction with drawings of other cities on the globe. Responses are "rotates toward east"; correct response to order of sunrises in various cities; correct responses (hours) for time zones.

5. Generalizing experience. Explain "one day of time" in terms of the earth, sun, rotation, and axis."

This example, as well as many others, are presented by Briggs to illustrate the application of instructional specification to a programed course in geography. It might interest you to know that the course was prepared by Dr. Vincent N. Campbell in collaboration with science teachers in local schools in the Palo Alto, California area (3). The materials were prepared for an experiment on program-directed study versus self-directed study (2).

The purpose of presenting the preceeding examples is to illustrate how an instructor, together with an instructional specialist and a media specialist, might predesign instructional conditions. Of course, there is more to it than what has even been mentioned above. Sequences of objectives must be reviewed in an overall fashion to make sure that media choices are reasonable. Specifications must be prepared for the production personnel. Informal tests of the feasibility of various sequences must be run on a small number of students whenever required. Only then may a prototype be developed and tried out.

If time allows, take an instructional objective you are familiar with and attempt to apply the steps presented above. Conduct an objective analysis on that objective to discover the sequence of enabling objectives. Attempt to identify the types of learning involved for each enabling objective and to specify the learning conditions for each type of learning. Then specify the instructional events and the medium for presenting each learning experience. Where you have questions or problems, note these and bring them up in the small group discussion period. If you think you have a better method of prescribing instructional conditions, please do not hesitate to mention this in your discussion groups as well.

References

1. Briggs, L.J., Campeau, P.L., Gagne, R.M. and May, M.A. Instructional Media: A Procedure for the Design of Multi-Media Instruction, A Critical Review of Research, and Suggestions for Future Research. Pittsburgh, Penn.: American Institute for Research, 1967. 176 pp.

2. Campbell, V.N. Self-Direction and Programed Instruction for Five Different Types of Learning Objectives. Psychology in the Schools, 1964, 1, 348-358.

3. Campbell, V.N. et.al. Water Unit. Palo Alto, California: American Institute for Research, 1964. (Mimeo).

4. Cotterman, T.E. Task Classification: An Approach to Partially Ordering Information on Human Learning. WADC TN 58-374, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, 1959. (AD 210 716).

5. Eckstrand, G.A. Current Status of the Technology of Training. Aerospace Medical Division. Wright-Patterson Air Force Base, Ohio, 1964. (AD 608 216).

6. Gagne, R.M. The Acquisition of Knowledge. Psychology Review, 1962, 69. 355-365.

7. Gagne, R.M. The Conditions of Learning. New York: Holt, Rinehart and Winston, Inc., 1965. 308 pp.

8. Gagne, R.M., Mayor, J.R., Garstens, H.L., and Paradise, N.E. Factors in Acquiring Knowledge of a Mathematical Task. Psychol. Monogr., 1962, 76, No. 7 (Whole No. 526).

9. Gagne, R.M. and Paradise, N.E. Abilities and Learning Sets in Knowledge Acquisition. Psychol. Monogr., 1961, 75, No. 14 (Whole No. 518).

10. Gustafson, H.W. Honsberger, W.D., and Michelson, S. Determination of Task Analysis Content. In WADD Techincal Report 60-593, Uses of Task Analysis in Deriving Training and Training Equipment Requirements. Wright Air Development Division, Wright-Patterson Air Force Base, Ohio, 1960. (AD 252 946).

11. Hatch, W.R. Approach to Teaching. Washington, D.C.: U.S. Government Printing Office, 1966. 36 pp.

12. Miller, E.E. A Classification of Learning Tasks in Conventional Language. Aerospace Medical Division, Wright-Patterson Air Force Base, Ohio. Prt. No. AMRL-TDR- 63-74. 1963 (AD 419 122).

13. Smith, R.G. The Design of Instructional Systems. Alexandria, Va.: The George Washington University, Human Resources Research Office, 1966. 85 pp.

14. Tracey, W.R., Flynn, E.B., Jr., and Legere, C.L.J. Systems Approach Gets Results. Training in Business and Industry, 1967, 4, 17-38.

Appendix A

Conditions for Principle Learning

"Conditions within the Learner • The prerequisite for acquiring the chains of concepts that constitute principles is knowing the concepts. Birds fly south in the winter is easily learned as a principle when the learner has already learned all four concepts involved in it. There is, of course, a kind of 'partial' learning of a principle that may result when the individual knows only some of the component concepts. Should a learner know all the concepts except south, it is apparent that some kind of principle could still be learned, but it would be an inadequate one.

As previously emphasized, knowing the concepts means being able to identify any members of the class they name. It is only when such prerequisite concepts have been mastered that a principle can be learned with full adequacy. Otherwise, there is the danger that the conceptual chain, or some parts of it, will become merely a verbal chain, without the full meaning that inheres in a well-established principle. It is unfortunately true that inadequate principles can be learned. It is a challenge for instruction to avoid these, and it is a challenge for measurement techniques to distinguish them from adequate ones.

Conditions in the Learning Situation • The major external conditions of principle learning are embodied in verbal instructions. The example of instructions used with round things roll will be useful to recall here.

1. The conditions of principle learning often begin with a statement of the general nature of the performance to be expected when learning is complete. In the previous example, the instructor says, 'I want you to answer the question, What kinds of things roll?' Why does he say this? Isn't he simply stating the principle, giving it away, so to speak? The main reason for making such a statement, which the learner 'holds in mind' during learning, appears to be this: It provides the learner with a means for obtaining immediate reinforcement when he has reached the terminal act. Having this statement for a model, he will be able to know when he has finished learning, and in many cases, when he has acquired the correct principle. Since principles may be long chains, the learner may need to have a conveniently retained reference to tell him when the end is

reached. The instructor, though, cannot be said to be 'telling the principle.' He doesn't state the principle itself, but only the kind of performance that will demonstrate the attainment of the principle.

2. Verbal instructions continue by invoking recall of the component concepts. The instructor says, 'You remember what roll means. . . . You remember what round means.' In many cases, the recall of these concepts is stimulated entirely by verbal means. In others (as in the example previously given) the class of stimuli that represent the concept may also be shown; the student may be asked to recall the roll event by identifying one, and a round thing by picking one out. Pictures, of course, may be used for this purpose as well.

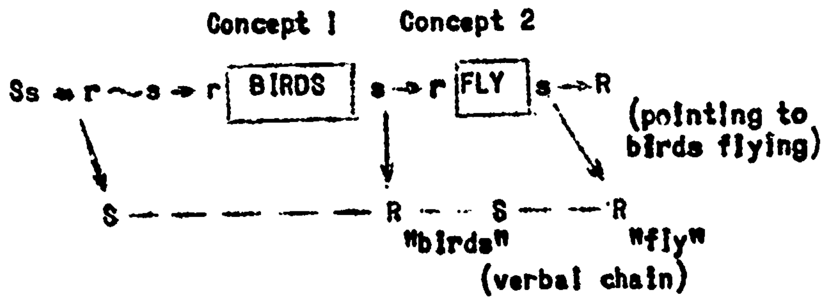
3. Verbal cues are next given for the principle as a whole. In our simple example, the verbal statement 'Round things roll' accomplishes this purpose. However, it should be noted that these verbal cues to the principle need not be an exact verbalization of the entire principle; they are in this case only because the principle is such a short one. If the principle were one from elementary geometry like 'An angle is formed by the intersection of two rays,' the verbal cues may be contained in such statements as 'Here are two rays. They intersect. We have an angle.' Such statements do not correspond exactly to an acceptable verbal definition. Yet they function as well or better in providing verbal cues to stimulate the learning of the principle.

4. Finally, a verbal question asks the student to demonstrate the principle. The instructor says, 'Show me.' The exact form is not of great importance so long as it truly requires the student to demonstrate the principle in its full sense. Added to this may be the requirement of asking the student to state the principle verbally, as when the instructor asks, 'What kinds of things roll?' But note particularly that such verbal statement is not essential to the learning of the principle, nor does it prove the student has learned the principle. Then why is it down? Probably for a very practical reason: the instructor wants the student to be able to talk about the principle later on, and so he teaches him the right words to say. This is undoubtedly useful, but it is important to note

that this kind of verbal chaining ('learning the definition') is an unessential part of principle learning itself.

5. The presence of some familiar learning conditions may be recognized in principle learning. Contiguity appears to be an important condition applicable to the time interval between the recall of component concepts (step 2) and the verbal cuing of the principle with these parts properly sequenced (step 3). Reinforcement is provided when the principle is exhibited in its complete form. The instructor may say, 'Right!' Even prior to this, the student may receive reinforcement by matching his terminal act with a form remembered from the initial instruction (step 1). Repetition has not been shown to be an important condition for this kind of learning, nor for its retention (Gagne and Bassler, 1963). However, the particular finding cited was obtained when the principles learned occupied an isolated position within other varieties of principles. The possible need for repetition in the form of review exercises to overcome the effects of interference should not be discounted. Forgetting, after all, is a highly likely occurrence with any kind of learned capability.

The capability acquired when a principle is learned must be somewhat abstractly represented, since by its very nature it cannot be related to specific stimuli, but only to the classes of stimuli that arouse its component concepts. As previously stated, a simple principle like 'birds fly' may be considered a chain of two concepts, and diagramed as follows:



The terminal response, here designated as 'pointing to birds flying,' may take many forms, so long as it constitutes a demonstration of the principle. Also indicated is the possibility of formation of the verbal chain 'birds fly,' which is the verbal statement of the principle, not an integral part of the principle itself " (7, p. 146-149).

IV - MEASUREMENT

Evaluation Sheet

We need to know how well the ideas and issues in this manual are communicated to you. You are the test audience for this material. To remove or strengthen the weak spots, to retain or improve the strong ones (if any), an account of your learning experience as you read these sections is crucial. Use this sheet as you study and jot down your reactions. Don't be concerned with typographical errors. We're concerned with how the message is coming through.

SECTION #________ TITLE__________________________

Give topic, paragraph or sentence, page	Importance Rate 1 2 3 4 5 Low - High	Clarity Rate 1 2 3 4 5 Low - High	Suggested Improvements

Over-all Rating of the Section: Use a 5-point Scale

Important					Understandable				
1	2	3	4	5	1	2	3	4	5
Very Low				Very High	Very Low				Very High

(Circle one) (Circle one)

PLEASE PUT ANY ADDITIONAL COMMENTS ON BACK

SECTION IV
Test Instruction

Table of Contents

SECTION IV

Measurement

H. Del Schalock

INTRODUCTION

Unlike the other topics covered in the Institute, measurement plays a central role in both instructional systems development and instructional systems research. With respect to instructional systems development, as Dr. Hamreus has pointed out, measurement serves three functions: 1) the determination of whether a learner achieves the level of performance that is expected of him, i.e., whether the terminal objectives and their various enabling objectives have been realized, 2) analysis of an instructional system to determine those aspects of it that are effective and those that are not, and 3) evaluation of the appropriateness of the conceptual framework that underlies the instructional system. This involves the re-analysis of the relevance and ordering of enabling analysis 2. From this point of view, measurement in instructional systems development is much more than simply "evaluating student performance;" it is an integral, inseparable, absolutely crucial part of the entire process. Because of this, the quality of instructional systems is closely related to the quality of measurement that is associated with them.

Measurement plays an equally critical role in instructional systems research.[1] It is also an expanded role. In instructional

---

[1] For purposes of the present paper the term instructional systems evaluation, or simply evaluation, will be used to refer to the various activities in the over-all task of instructional systems development which depend upon measurement. The term instructional systems research, or simply instructional research, will be used to refer to a set of activities that are considerably different than those involved in evaluation. Evaluation is a relatively limited, narrow range of activities that lead to information which has relevance only to a specific instructional system within a specific context. It does not lead to principles, generalizations, laws, or theory. Research, on the other hand, refers to the much broader class of activities that lead to the extension of generalizable knowledge. Research that relates to instructional systems focuses upon such matters as the appropriateness of an instructional system for various kinds of children, the relative effectiveness of alternative media forms or presentation strategies within a given instructional system, the generalizability of an instructional system across subject matter areas, etc.

systems development, measurement focuses only upon the objectives of instruction; in instructional research, it maintains this focus, but includes two others as well. In their most general terms, these are: 1) descriptions of the <u>treatment conditions</u> employed in instruction, i.e., the <u>content</u> of the system, the form of the <u>media</u> in which it appears, and the methods by which it is presented, and 2) descriptions of the <u>contingency factors within the setting which interact with the treatment conditions</u>. These generally are of three kinds: a) learner characteristics, e.g., age, intelligence, background of experience, b) teacher characteristics, e.g., background of knowledge, training, interests, and c) setting characteristics, e.g., number of children in a class, availability of materials, time of the day. It is these factors, of course, which determine whether an instructional system will be effective or not effective, and as such they must be controlled as far as possible in all instructional research. In fact, without the measurement and control of these factors, instructional research would not be possible - just as instructional evaluation would not be possible without the assessment of instructional objectives.

Thus it is that both instructional evaluation and instructional research are completely dependent upon measurement. Without it, there is neither. But the issue here is not one of whether there is or is not measurement; most educators know that measurement has to be a part of these activities. The real issue is whether there is "good" or "poor" measurement, and most educators are not able to make this distinction. Broadly speaking, the purpose of the present chapter is to help them be able to do so. Specifically, the intent of the chapter is to: a) sensitize the reader to the critical role of measurement in evaluation and research, b) provide for him the information needed to wisely select the formal, existing measures that he wishes to use in his evaluation and research programs, and c) to develop in him the competencies needed to build the high quality informal, "teacher-made" measures that he needs to use in these programs. Developing the competencies needed to build formal, standardized measures is beyond the pale of the present effort.

Hopefully, the chapter also will clarify the nature of educational and psychological measurement, for by-and-large it is badly misunderstood. While it is not hard to understand or accept the principles involved in some of the more common measurements used in the natural sciences, for example, length, weight, and volume, it is hard sometimes to understand and accept the

fact that the measurement of such characteristics of individuals as achievement, intelligence, aggressiveness, and anxiety involves basically and essentially the same thinking and the same general procedures. It is also hard to accept the fact that if the rules of measurement can be set up on some rational or empirical basis, measurement of anything is theoretically possible. From the point of view of the writer it is important that educators somehow develop this point of view for with it they will not reject the possibility of measuring some property simply because it is complex and elusive. They will, as Guion has put it " . . . understand that measurement is a game that they may or may not be able to play with this or that property at this time. But they will never reject the possibility of playing the game, though they may realistically understand its difficulties". (1966, p).

A POINT OF VIEW TOWARD MEASUREMENT IN EDUCATION

In addition to recognizing the place of measurement in instructional evaluation and research, and being committed to the idea that that which needs to be measured can be, the educator must also be aware of some of the realities that are involved in educational measurement. In this writer's opinion four such "realities" are paramount.

1) An educator or educational researcher must understand the fundamental nature of measurement, particularly as it applies in the behavioral sciences, if he is to pursue evaluation and research activities judiciously. Simplified guides as to when to use what kinds of measures upon whom are not enough. If education is to be effective, and if a "science" of education is to emerge, the nature of educational measurement must be understood by those who use it. Lindquist states the view well when he says "If measurement is to continue to play an increasingly important role in education, measurement workers must be much more than technicians. Unless their efforts are directed by a sound educational philosophy, unless they accept and welcome a greater share of responsibility for the selection and clarification of educational objectives, unless they show much more concern with what they measure as well as how they measure it, much of their work will prove futile and ineffective." (1950, p 158).

Toward this end, an Appendix is attached to the present chapter (see Appendix A) which introduces the reader to the concept of measurement generally and to the characteristics of measurement in the behavioral sciences specifically. While it is not assumed that all will read this material during the course of the institute, the reader is encouraged to become familiar with it as time permits.

2) An educator or educational researcher must be willing to undertake the measurement of classes of educational objectives that heretofore have gone largely unmeasured. Thus far in education both standardized achievement tests and informal teacher examinations have focused primarily upon the measurement of content objectives that derive from established school subjects. Objectives concerned with such things as sensitivity or consideration toward another, artistic abilities, artistic aesthetic pre-

ferences, moral values, attitudes toward social institutions and practices, habits relating to personal hygiene ar . physical fitness, managerial or leadership ability, etc. have been seriously neglected--even though most educators hold them to be of importance. The point of view adopted here is that if these are important classes of objectives, they should be so specified and instructional systems developed to bring them about. In line with earlier comments, this would require that appropriate measures of the[illegible] jectives be established.

3) If the educator or educational researcher does undertake the measurement of the full range of educational objectives to which he is committed, he will have to rely upon more than "paper and pencil" measures. This point of view rests upon the assumption that the aim of any educational achievement measure is to obtain a sample of the criterion behavior, i.e., the behavior which is specified in the behavioral objective, or a sample of behavior or the products of behavior which presumably relate to the criterion behavior. How else can one obtain evidence that an educational objective has in fact been realized? Given this point of view, paper and pencil tests are not at all appropriate to the measurement of the full range of educational objectives. When the focus of measurement is upon "knowledge", e.g., concepts, facts, and principles, or upon the application of knowledge to a set of tasks that require only the manipulation of symbols, e.g., the solution of a mathematics problem, outlining the steps involved in building a house, or writing a theme, on the expression of consideration in one's relations with others, paper and pencil measures are perfectly appropriate. However, when the focus of measurement is upon building a house, or upon relating considerately to others, a paper and pencil measure won't do--unless one is willing to make the assumption that being able to outline how to build a house or how to behave considerately is in fact related to the ability to build a house or to act considerately. While most people would probably accept the idea that the knowledge factor is related to the concrete performance factor, few would accept the idea that the relationship is perfect. If this is true, then measures other than or in addition to those which require only the manipulation of symbols are needed in order to assess some of the objectives of education.

Toward this end it is proposed that the educator or educational researcher must be familiar with the full range of mea-

sures available to the behavioral scientist. These include such "obtrusive" measures as interviews, standardized objective tests, teacher-made tests, systematic observation (face-to-face observation, tape recordings, video-tape recordings) and standardized projective tests, and such "unobtrusive" measures as physical traces through erosion and accretion, e.g., wear on library books or the accumulation of used paint tubes, documents and products, simple observations, and contrived observations through hidden hardware. Each of these are specialized measurement methodologies, some of which are appropriate to one kind of measurement problem and some appropriate to others, but in order to handle the full range of measurement tasks that he faces an educator or educational researcher needs to be competent in them all.

Limitations of space and time make it impractical, however, to enter into a discussion of the various methodologies in the present paper. They are complex and each has its own unique set of problems. To provide the reader with some idea as to the nature of these measures, however; the kind of data that they provide, and some of the particular strengths and weaknesses of each, the major classes of measurement methodologies are summarized in Appendix B. While this brief summary will in no way prepare the reader to <u>use</u> the various classes of measures hopefully it will help sensitize him to the <u>possibility of their use</u>. Excellent discussions of the operations involved in the various measures appear in Festinger and Katz (1953), Lindzey (1954), Mussen (1960), Gage (1963), Kerlinger (1965), and Webb, et al. (1966).

4) <u>Because of the "inexact" nature of measurement in the behavioral sciences (see Appendix A) educators or educational researchers should bring several different classes of measures to bear upon each behavioral objective that is to be assessed.</u> There are two reasons which underlie this recommendation: 1) any single measure can hope to assess only one or a few of the many enabling objectives which comprise a terminal objective, (or perhaps even only a few of the properties which define <u>an</u> enabling objective!) and these only within a limited sample of the situations in which the enabling objective is reflected, and 2) because all measures, whether in the physical or behavioral sciences, involve a degree of error. This is especially the case in education, however, where many of the measures must of necessity be indirect measures, i.e., they are <u>indicants of or supposedly related in some way to</u> a given educational outcome. Given

these two conditions the educator and educational researcher need to employ what Campbell and Fiske (1959) have called a "multiple operations" approach to measurement. In brief, it calls for multiple measures to be used in every study, and assumes that in combination the measures will "share in the relevant components (of that which is to be measured) but have different patterns of irrelevant components" (Webb et al., p. 3). The basic assumption underlying the procedure is, simply, that once a proposition has been confirmed by two or more independent measurement processes, the uncertainty of its interpretation is greatly reduced. Campbell and Fiske argue that the most persuasive evidence for the existence of that which is being measured comes through a triangulation of measurement processes. The basic assumption underlying the approach is that if a proposition can survive the onslaught of a series of imperfect measures, with all their irrelevant error, confidence can be placed in it.

A sample listing of multiple criterion measures by class of educational outcome appears as Appendix C.

MEASUREMENT IN INSTRUCTIONAL SYSTEMS DEVELOPMENT

While a rather wide range of standardized achievement measures exist (see Buros, 1965), most measures that are used in the evaluation of instructional systems are of the informal, teacher-made variety. Considering the almost limitless number of educational objectives that are open to selection, and the part that individual teachers play in their selection, this probably will continue to be the case for many years to come. Since these measures are nearly always specific to each instructional system, and therefore need to be developed by the person who is developing the system, the focus of the present section of the paper is upon the development of this kind of measure.

Within this focus five topics are considered: 1) planning the measure, 2) developing the measure, 3) trying out the measure, 4) evaluating the measure, and 5) the "formal" attributes to be considered in judging the worth of an informal measure. The procedures and attributes reviewed within these topics are applicable generally across all classes of measures within the behavioral sciences (see Appendix B), that is, they are not limited only to paper and pencil measures. However, for persons wishing concrete, specific guidance in the development of paper and pencil classroom tests five recently published books provide excellent references: Sax, G. The Construction and Analysis of Educational and Psychological Tests: A Laboratory Manual (1962); Ahmann, J.S. and Grock, M.D., Evaluating Pupil Growth (1963); Stanley, J.C. Measurement in Today's Schools (1964); Gronlund, N.E. Measurement and Evaluation in Teaching (1965); and Ebel, R.L. Measuring Educational Achievement (1965). The pros and cons of various item types, including essay items, how these can best be developed, how to administer, score and use tests, etc. are discussed in detail. An article by Eva Baker on the development of prototype items is also of value in this respect (see Institute Materials).

Planning the Measure

As Dr. Hamreus has indicated, one should begin the development of the measure to be used in assessing criterion performance as soon as the hierarchy of enabling objectives has been identified. There are two reasons for this: 1) it eliminates the pitfall of assessing "that which has been taught" in contrast to "that which

was supposed to have been taught", i.e., the behavioral objectives,[2] and 2) in the process of building a measure to assess the various enabling objectives clarification may come in the enabling objectives themselves. Given a hierarchy of enabling objectives there are three tasks to be accomplished in the test planning phase: 1) a decision as to the fidelity of the items to be used in the measure, that is, the extent to which the response to items within the measure will require the performance of the concrete behavior specified in the behavioral objective, in contrast to the performance of behaviors which are only related to it, e.g., building a house vs. describing how one would be built or acting considerately vs. describing what acts of consideration involve[3], 2) a decision as to the representativeness of the items to be used in the measure, that is, the number and range of situations to be sampled by the items in which the behavior is reflected, e.g., does one ask to have one house of one kind built or three houses of one kind built, or one house of three different kinds built (to test for transfer?), and 3) the weight to be given to each item. The question of fidelity, of course, revolves around the issue of inference: is one willing to accept as evidence of the accomplishment of a terminal objective something less than the actual performance of that objective? Is one willing to accept a behavior supposedly related to the terminal objective as evidence of the objective, and if so how close or how distal can the relationship be and still be acceptable? For example, is one willing to accept as evidence of considerateness toward parents a verbal statement to the effect that the examinee is considerate? Would observation of considerateness toward animals or children be any better? Would considerateness toward peers or

---

[2]To eliminate a reverse kind of pitfall, there is a generally agreed upon rule in educational measurement that it is not cricket to focus an instructional system upon an item that is included in the evaluation measure.

[3]It may be of course that the "behavioral objective" itself involves only the description of how to build a house. In this sense, the issue of fidelity rests initially with the instructional systems builder. Even so, it is also always an issue in the development of measures to assess a behavioral objective, for the circumstances of an educational setting may force the test developer to settle for a level of fidelity that is less than desired.

a teacher be better still? These are the kinds of issues involved in the question of fidelity, and they are crucial in the process of achievement measurement.

The question of representativeness revolves around the issue of generalizability: how many and what kind of situations must be sampled or observed in order to feel confident that the criterion behavior is sufficiently well established that it is a functional part of the behavioral repertoire of the examinee? The question of item weighting is simply a matter of deciding whether some items will be more informative of criterion performance than others.

Unfortunately, there are no firmly established rules to guide decisions on any of these three questions. The level of fidelity desired in an instrument is a matter of personal preference (though one probably should be able to defend it) interacting with the reality demands of a situation. It is sometimes difficult to set up a testing situation which has architectural students building houses. The extent of evidence required as to the representativeness of the performance measure is also largely a matter of personal preference and reality demands: generally speaking, the broader the sampling of items the better the evidence, but also the higher the cost. Establishing rules for weighting item responses is also arbitrary, though in most short answer tests of educational achievement equal weight is given to all items. In measures involving essay questions, or products, or interpersonal behavior the matter of assigning item weights is equally arbitrary. The single rule that could be thought to operate in all these decisions is that of logic: the fidelity of items and the representativeness of situations sampled and the assigning of item weights must somehow meet the demands of elementary logic. If a decision on any of these matters wrenches one's credulity, it is likely that even this single criterion has been violated.

The lack of established rules for making planning level decisions should not be taken to mean that these are insignificant decisions. Indeed, of all the decisions made in relation to measurement they are perhaps the most basic, for all else stems from them. In order that the reader get a firm grasp of the concepts involved in these decisions the topics of fidelity and representativeness are reviewed again under the heading Formal Attributes of Informal Measures.

Developing the Measure

Given the enabling objectives that are to be measured, and clarity as to the fidelity of items and the range of situations to be sampled by those items, one is then ready to proceed with the development of the measure. Three steps are involved in this process: 1) the specification of the set of operations that are to be used in assessing the status or condition of each of the objectives that are to be measured, (the development of an "item" pool), 2) the specification of the conditions under which these operations are to be made (the development of a set of directions) and 3) the specification of the rules by which numerals are to be assigned to these operations (the development of a "scoring" system). In deciding upon the set of operations to use (the "item types") it is necessary to look to the full range of measurement methodologies available to the behavioral scientist (see Appendix B). In so doing the aim is to find the methodology that provides the most appropriate set of operations for the assessment of a given objective at a given level of fidelity. Thus, if the objective being measured is considerateness toward others, and the level of fidelity is set at "identical elements", i.e., it requires the performance of the concrete behaviors specified in the objective, the measurement methodology most appropriate to the task would probably be systematic observation. Other appropriate methodologies would include the interview, simple observation and contrived observation. Perhaps the least appropriate methodology would be that involving teacher-made paper and pencil tests! Examples of "item types" when using systematic observational procedures might include the recording of all instances of helping another when help was sought, helping another when help was not sought, "surprising" another with unexpected gifts, thoughtfulness, etc.

The selection of a methodology, and the various "item types" within it, represents the first important consideration in item development. Once this is done it can be combined with the specification of situations to be sampled by the items, e.g., the considerateness of the examinee at home with her mother, her younger brother and her father and her considerateness at school with her "three best friends", and the development of items can get underway. Generally speaking, a measure should contain as many items as is practically possible, for the wider the sample of items within situations the more representative the responses to the measure should be. Ordinarily, it is wise to develop a larger pool of items than ultimately will be needed.

In combination, the selection of an appropriate measurement methodology and the selection of appropriate item types within it provide a basis for assessing the relevance of a measure to that which is to be measured. (For those familiar with test theory, the concept of relevance is comparable to the earlier concepts of face or content validity--see Standards, 1966). As was the case with planning decisions, the single rule that may be thought of as operating in development decisions is that of logic: if a decision as to measurement methodology or item type stretches one's credulity as they pertain to a particular objective then it is likely that the criterion of logic has been violated.

Since paper and pencil tests are widely used by educators, and are appropriate for the measurement of many of the educational objectives in the cognitive domain, special attention will now be given their development. In entering this discussion it is to be realized that all of the planning and development decisions that have been reviewed thus far apply to the development of paper and pencil measures.

As with any other methodology it is wise in developing paper and pencil measures to develop a larger pool of items than ultimately will be needed. Also, it is wise to employ more than one type of item, e.g., completion, true-false, multiple choice, ranking. Items should be phrased so that the content of the statement rather than its form will determine the answer, that is, avoid telltale words such as "always", "never", "entirely", or "absolutely", which usually lead to an answer being false, or such words as "may", "sometimes", and "as a rule" which are most often associated with true statements. All items of a particular type should be placed together in the test.

A particularly critical issue for the builder of classroom paper and pencil tests is that of item difficulty. Obviously, if an item is so easy that everyone taking the test answers it correctly, then it is of no value in discriminating between those who have more and those who have less of the property being measured. The same problem exists if an item is so hard that no one answers it correctly. Yet for purposes of morale, the reduction of test anxiety, etc. it may be good policy to put in a few items at the beginning of a test that are extremely easy, and for purposes of maximum discriminatory power it may be desireable to put in a few items at the end of the test that few are

likely to get right. As yet there are no agreed-upon rules to govern treatment of this dilemma, but a commonly appearing rule of thumb suggests that most items in the test should be of approximately 50 percent difficulty, that is, approximately half of the group being tested should know the answer. This rule evolves from experimental work which shows that under these conditions a test has maximum discriminative power. In practice, however, test makers generally produce test items having a wide range of difficulty with only the average level of difficulty being approximately 50 (Guion, 1966, p. 200).

In advance of giving a test one can never be sure, of course, how difficult a particular item actually is. A procedure for determining item difficulty is outlined in the section on evaluating a measure.

After the item pool has been established, the two remaining instrument development tasks can be undertaken, namely, the specification of the conditions under which the items are to be administered and the specification of the rules by which numerals are to be assigned to that which is assessed by an item. In the example where measurement involved the observation of a girl relating to others, specification of the conditions under which the items are to be administered might include such statements as "after dinner alone with her mother in the kitchen" or "at bedtime with her brother upstairs." Directions for taking a paper and pencil test also must be clearly stated. In general, instructions for taking a test should be so clear that the least able student in the class knows what he is expected to do, even though he can't do it.

Rules for assigning numerals to that which is being assessed ("scoring" rules) vary widely with measurement methodology, item type, and weighting decisions. For example, when preconceived category sets are used in systematic observation procedures the scoring rule is usually 1 or 0, that is, a category appeared or it didn't. When rating scales are used the scoring rules may be something to the effect that "when the behavior in question occurs frequently, check scale position 3, when it occurs seldom or not at all check position 1; when it occurs at a frequency somewhere between 3 and 1, check position 2." Most short answer paper and pencil tests also provide a score of 1 or 0, although some involve a +1 or a -1 with the -1 serving as a penalty for guessing. Formulas do exist that permit correction of "chance" or "guessing" on short-answer paper and pencil measures (Guion, 1966).

Trying out the Measure [4]

Since it is impossible in advance to know how good a measure is, or to know which items are good and which are poor, a tryout of the measure is essential. Two rules need to govern the tryout: 1) every reasonable precaution should be taken to insure the best of measurement conditions, and 2) the time allowance for obtaining the measure should be generous. Since both rules mean somewhat different things for different measurement methodologies, e.g., in systematic observation the setting being observed needs to be "natural", attention is to be directed to the normal course of events, and observer influence needs to be at a minimum and in paper and pencil testing the situation needs to be quiet, attention has to be directed to a specific task and teacher influence (presence) needs to be high, administration and timing rules will not be discussed here. For those interested in administration, timing, etc. in relation to most of the methodologies in the behavioral sciences see Festinger and Katz (1953), Lindzey (1954), Mussen (1960), and Kerlinger (1965). For those interested in the administration of paper and pencil measures in the classroom see Stanley (1964) and Ebel (1965).

A good deal can be learned about a measure from simply attending to the response of examinees to it. For example, in an interview situation one can obtain a fair idea as to which items are clear and which are confusing, or which are probing "sensitive" areas and which are not simply by observing the respondent's reaction to the items. Similarly with a paper and pencil measure: the overt reactions of students to the test will often suggest whether it is understandable or not, whether it is hard or easy, whether it is relevant to the objectives of the class or irrelevant to them, whether there is too much or too little time, etc. While a measure will not stand or fall upon information as informally obtained as this, it is nevertheless information that is worth obtaining.

---

[4] Trying out a measure before it is to be used for evaluation purposes is based upon the assumption that a measure is to be developed that will be used more than once. If this is not the case the test should be read or the measure tried by at least one other person before it is administered formally.

Evaluating the Measure

In addition to the informal assessment of a measure, as described above, a formal assessment may be made once the measure has been administered to a number of people that are representa- of the target audience specified for the measure. The nature of the evaluation depends, however, upon the use to which the measure is to be put. For example, if a test is to determine mastery to an absolute criterion and the criterion involves passing all items, then one looks only at the number of items answered correctly; if it is to serve a diagnostic function, then interest is not so much in how many but in which items were passed or failed. If a test is to distinguish between students on that which is being measured, such as academic achievement, intelligence, etc., then one looks at each test item in terms of its ability to do so. Since measures taken in the classroom most commonly serve this last function, attention will be directed in the present paper to it. For reasons of space, the discussion will take as a point of focus only a short-item paper and pencil measure of achievement. The discussion can be generalized, however, to the evaluation of all measurement methodologies.

In evaluating a test for its effectiveness in discriminating between students, the range and distribution of the total test score is only a rough indication of its adequacy. Since a test is made up of items, a total test score is dependent upon the discriminating power of those items; if only part of the items are discriminating then the test score is reflecting the work of only those items. It is possible, through item analysis procedures, to determine which items are and are not discriminating. The value of an item analysis is that it permits one to know which items to eliminate in the next edition of the test and how many new items to add. (Obviously, in using this procedure, an item analysis has to be made each time new items are added). By following item analysis procedures one can continuously in-

---

[5] A teacher must also decide whether to evaluate students in terms of an absolute criterion, that is, whether to evaluate students against a standard which the teacher or someone else sets or to evaluate them in terms of their performance relative to the performance of others. A recent paper by Gerlack, Schutz, and Baker (see Institute Materials) provides an excellent discussion of this issue.

crease the power of a test to discriminate, as well as increase one's confidence that individual items do in fact discriminate across various class groups.

The simplest procedure for testing the discriminating power of an item is to determine the number of correct responses to the item by the students who rank in the highest 27 percent of the class on the test as a whole, and to compare this with the corresponding number in the lowest 27 percent of the class. The items for which the number of correct responses of the high group most exceeds that of the low group are most discriminating; those in which the number of correct responses of the high group falls behind that of the low group, and those in which the numbers are the same, are not discriminating. These are the items that should be rewritten or discarded. The basic assumption underlying this procedure is that the initial total test score is a relatively accurate measure of the behavior under consideration for all items subsequently kept and added are based upon that score. To many, this is a difficult assumption to accept.

Formal Attributes of Informal Measures

In the preceeding sections of the paper attention was directed to three concepts: fidelity, representativeness and relevance. While these properties or attributes of a test were dealt with in some detail within the context of test development, they are of sufficient importance to be reviewed a second time. Also, to be understood fully, they need to be seen within the context of other test properties, for example, reliability, validity, accuracy, etc. The purpose of the present section of the paper is to set these various concepts in this kind of perspective.

A basic assumption underlying the present discussion is that measures of educational objectives or educational achievement are direct measures, that is, they assess specified behavioral objectives, and in this sense they differ from many other behavioral science measures that are of necessity indirect, e.g., measures of intelligence, anxiety, authoritarianism, etc. In the case of assessing educational objectives the task is to obtain measures of the behavior in question, or measures which are believed to be related to it; in the case of assessing such characteristics as intelligence or anxiety, which in themselves are inferred characteristics or constructs (see Appendix A)

the task is to obtain measures of indicants of the characteristic or construct. While the concepts of relevance, fidelity, representativeness, reliability and accuracy are appropriate to both direct and indirect measures, it is only in the arena of indirect measurement that the notions of operational definitions and construct validity come to bear. The concept of predictive validity is also applicable to both direct and indirect measures, but it becomes an issue only when prediction of some form or another is involved (see Standards, 1966). Both the concept of construct validity and predictive validity will be discussed in the section on MEASUREMENT IN INSTRUCTIONAL SYSTEMS RESEARCH.

The discussion which follows has as its focus only the attributes of direct measures, in this case the attributes of measures which are aimed at the assessment of educational outcomes. The intent of the discussion is to provide a summary of the attributes needed by measures of this kind before they can be used with confidence.

Relevance. As indicated earlier, the relevance of a measure refers to the extent to which it is a logically appropriate measure for that which is to be measured. Technically, the relevance of a measure is determined by two considerations, the selection of an appropriate measurement methodology and the selection of appropriate item types to be used within that methodology. If either selection stretches credulity, then the relevance of the test is suspect. As used here, the concept of relevance is intended to replace the earlier concepts of face validity and content validity (see Standards, 1966). In terms of the concept of error variance (see Appendix A) relevance is one means of reducing constant error.

Representativeness. The representativeness of a measure refers to the extent to which it samples the situations in which the behavior under consideration is reflected. Generally speaking, the greater the number of situations sampled the more confidence one can have in judging whether the behavior observed is or is not a functional part of the behavioral repertoire of the individual. As in the case of relevance, the primary means for determining the representativeness of a measure is through logical analysis: if the situations sampled are not representative of the full range of situations available for sampling then the measure is suspect. A formal means for estimating representativeness also exists; namely, the comparison of scores obtained on

two forms of the measure. If scores are similar on both forms then it may be assumed (granting the acceptability of the measure by logical analysis) that the sampling of situations is adequate. The two forms of the measure may be developed separately or they may be formed simply by randomly dividing the situations sampled into two equal halves. This procedure will be recognized as "alternate form" and "split-half" reliability measurement, but as used here the concept of represntativeness is intended to replace these earlier concepts. In terms of the concept of error variance representativeness is also a means of reducing constant error.

Fidelity. As used in the present paper the fidelity of a measure refers to the extent to which a measure calls for the performance of the concrete behavior specified by a behavioral objective; it may require behavior that is isomorphic to the objective (identical elements) or it may call for behavior that is only in some way related to it. Ideally, all measures of educational objectives should be isomorphic, but since the realities of educational settings sometimes make this impractical the general rule to be followed is to make one's measures as high in fidelity as is practically possible.

Many years ago Lindquist (1950) proposed four alternative approaches to the measurement of educational objectives: 1) give the examinee occasion to do some of the things that are specified by the objective (an isomorphic situational response test); 2) give the examinee occasion to do things similar to some of those specified by the objective, (a "related behavior" situational response test); 3) describe a situation in which the examinee would have occasion to do what the objective specifies, and then ask him to tell what he would do in this situation or how he would do it (a "verbalized behavior" situational response test); and 4) discover whether or not the examinee knows the facts, rules, principles, etc., that are presumably essential or conducive to the desired behavior, (a "knowledge" test) (p. 146). At the time Lindquist wrote his paper educators were measuring essentially at the knowledge level, and his plea was to get them to move to a higher level of fidelity. Today they are still measuring essentially at the knowledge level, and the plea is still relevant. This is not meant to imply that measurement should avoid focusing at the knowledge level; indeed, many educational objectives are focused entirely at that level. Also, there is obvious truth in the argument that knowledge is

essential or conducive to the overt behavior with which an ultimate educational objective is concerned, i.e. there is a relationship between what and how much an individual knows and how he will behave in certain situations. The point of the discussion here, however, is that while the measurement of knowledge is in many instances a worthwhile goal, it cannot substitute for the situational response type measures as measures of educational objectives.

To date the concept of the fidelity of a measure has not appeared in the literature on test theory, so the concept has had little empirical testing. On the surface, however, it appears to be a useful concept and so it has been included here. In terms of the concept of error variance it too contributes to the reduction of constant error.

Taken together, the three measures of relevance, representativeness and fidelity in <u>direct</u> measurement take the place of the concept of validity in <u>indirect</u> measurement.

<u>Reliability.</u> Reliability is that attribute of a test which speaks to the consistency with which it measures that which it measures. As discussed earlier all measurement, whether in the physical or the behavioral sciences, contains a certain amount of chance or random error. Two sets of measurements of the same characteristics or properties of the same individuals will never exactly duplicate each other. This is termed the <u>unreliability</u> of measurement. At the same time, however, repeated measurements of a property within an individual will (if the measure is at all appropriate), show <u>some consistency</u>. For example, if a boy was the best reader in the room the first time that a class was tested it is highly probable that he will be among the best readers on another testing, even though he may not be the best. This tendency toward consistency for a repeated set of measures is what is known geneally as reliability. Technically, reliability may be defined as the extent to which a set of measurements is free from variance due to random error, but in fact it tends to be more a measure of <u>the consistency of the individual that is being measured</u> than it is a measure of the test per se.

A number of factors may contribute to a lowering of the reliability of an instrument: a) response variation by the subject due to fatigue, illness, an "incorrect set", etc., b) variation in administration or the administrator of the test,

c) variations in scoring, etc. As long as the sources of error remain unsystematic, however, i.e., not constant, their threat to measurement is not great.

Within this framework there is essentially only one procedure for obtaining an estimate of the reliability of a measure, namely, retesting an individual or set of individuals with the same test. While there are a number of limitations inherent in test-retest reliability, for example if one waits very long between measures individuals will change so much that the repeated measure will of necessity be quite different from the first, or if one retests too quickly the second score will be subject to recall error from the first, it is still the one measure that provides data relevant to the concept of reliability. Reliability estimates are needed for all measures, whether direct or indirect, for random error always enters measurement.

A special problem is encountered in reliability estimation when the measurement methodology being used does not call for the rigid control of test stimulus materials, e.g., when the measure is the description (categorization) of the free play of children, teacher-child or parent-child interaction, group problem solving behavior, etc. Here the stimulus situation is never the same twice (unless one is working in a controlled, experimental situation) and consequently there is no reason to assume that the behavior in question will ever be the same twice. Nevertheless, it is still possible to obtain an estimate of the consistency of these kinds of measures if one is interested in doing so. The only requirement is that the situation that is being observed be as similar as possible on repeated measurement occasions, e.g., observing the same children in the same play area at the same time of the day under as many of the same conditions as possible.

Accuracy. The accuracy of a measure has often been confused with reliability, but they are in fact two quite different concepts. A test may be quite reliable (a person may consistently receive a similar score upon repeated measures) but it may not be accurate, that is, the test could show him with x amount of a characteristic when in fact he has an x + 1 amount. Nor is accuracy comparable to relevance, representativeness, fidelity or validity. A test may be valid etc., that is, it tests x when it is supposed to test x, but it may be so gross as not to be able to distinquish between x and x + 1.

From a technical point of view, the concept of accuracy has no legitimate meaning in the behavioral sciences. This is because these sciences have access only to nominal and ordinal level data (see Appendix A); one must have an absolute measure against which to judge (such as those which exist in the Bureau of Standards) in order to have a measure of accuracy. This first of all requires measurement at the ratio level.

Practicality. While practical considerations can never justify the use of a test which gives worthless information, a relevant and technically sound test cannot be used where it is impractical. Thus users of tests need always to seek a viable "trade-off" between the demands of good measurement and the demands of reality. The major factors which need to be considered generally in relation to practical matters are cost, time required for administration and scoring, ease of administration and scoring, the availability of comparable test forms, user acceptability and potential usefulness of results. Nearly all textbooks in educational and psychological measurement deal with these topics, so they will not be pursued here.

MEASUREMENT IN INSTRUCTIONAL SYSTEMS RESEARCH

As indicated earlier a rather sharp distinction can be made between instructional systems evaluation and instructional systems research: instructional evaluation pertains to that class of activities that leads to the acceptance, modification or rejection of a specific instructional system within a specific context; instructional research involves that class of activities that leads to the extension of generalizable knowledge. Evaluation involves essentially three functions: 1) the determination of whether a learner achieves the level of performance that is expected of him, 2) an analysis of the instructional system being used to determine those aspects of it that are effective and those that are not, and 3) an analysis of the appropriateness of the conceptual framework that underlies the instructional system in light of the empirical evidence that derives from analysis 2. In this sense instructional systems evaluation is a continuous, recycling process that leads to an increasingly clear statement of terminal objectives, their enabling objectives, and the instructional system that leads to their development.

Instructional research involves quite a different set of functions and a much broader focus. In contrast to the development-testing-development cycle, in instructional evaluation, with its focus upon <u>products</u>, research involves a hypothesis-testing-hypothesis cycle with a focus upon <u>facts</u> or <u>principles.</u> A "package" or "a usable product" evolves from an evaluation activity, and it may or may not be generalizable, "words" evolve from a research activity and are generalizable. Also, in contrast to the purely descriptive-analytic nature of evaluation, the strategy of research is to compare, correlate and predict. Its aim is to test such matters as the appropriateness of an instructional system for various kinds of children or the effect of using the system under various kinds of conditions and with various kinds of teachers. As such, it requires the introduction of experimental manipulation and control (research design) and the use of analysis techniques which facilitate the drawing of inferences from observations (statistical analyses). It also requires a broader measurement focus. Instead of being concerned only with measures of educational achievement or outcome, as is the case in evaluation, instructional research of necessity focuses upon three classes of measures: 1) the treatment conditions

that comprise the instructional system (content, media, methods), 2) the contingency factors which interact with the treatment conditions to make them effective or ineffective (learner characteristics, teacher characteristics, setting characteristics), and 3) the outcome or achievement measures. Since the measurement of educational achievement has been dealt with in the preceding section of the paper the present section will focus only upon treatment and contingency measures.

The aim of this section of the paper is threefold: 1) to describe the nature of treatment and contingency measures, 2) point up the similarities and differences between these measures and achievement measures, and 3) to introduce the reader to O.K. Buros' Mental Measurements Yearbooks as a guide to the published measures that are available to the educator and educational researcher. Because of time and space limitations these comments must be brief.

The Nature of Treatment Measures

Treatment measures involve essentially detailed descriptions of that which comprises an instructional system. This involves three separate components: 1) the content of the system, e.g., the firepower of a nuclear submarine or the theme of a Wagneriar pera, 2) the media used to transmit the content, e.g., the printed page, pictures, models, records, television, films, real-life observation, and 3) the methods used in dealing with the media and its content, e.g., individual study followed by an examination, lecture, discussion which follows an "inquiry" model, small group discussion, teaching machines. Historically, discriptions of these components have been extremely gross--much like they appear above. Within recent years, however, more detailed descriptions of these components are beginning to find their way into the research literature, and as a consequence the nature of the data on the relationship of treatment conditions to learner outcomes is starting to take a different shape (cf Taba, 1964, Bellack, 1965).

The detailed description of content, media and method provides a set of measurement problems that are essentially the same as those that confront the developer of educational achievement measures. As a beginning step he must find or develop a framework which lets him know what he is to measure. If he has

to develop the framework he is doing essentially the same thing that the instructional systems designer does when he breaks out the enabling objectives that are in service to a terminal objective. If the focus is upon content, the task is to find a framework which permits a meaningful ordering of the substantive information within a unit of subject matter. Depending upon the particular focus of the research this could follow the scheme of Gagne' (1964) and study content from the point of view of structure i.e., the hierarchy of discriminations, associations, concepts, principles, etc. that are involved in it, or it could be approached from the point of view of a topical analysis, i.e. the substantive topics covered. If the focus is upon method, the task is to find a framework which permits a meaningful ordering of the various aspects of a teacher's behavior that are designed to facilitate the learning process. Detailed analysis schemes of this kind are now appearing, but as yet there is some question as to their utility (Schalock, 1967). If the focus is upon media the task is to find a framework which permits the classification of various media forms. Once a framework has been developed which identifies that which is to be measured (described) the process of developing the respective measures can begin.

By-and-large, the development of treatment measures follows the same patterns and involves the same set of operations as does the development of outcome measures. Both involve direct measurement, and as such attention must be given to the issues of relevance, representativeness, fidelity and reliability. Since measures of treatment conditions have not as yet been developed, the measurement task of the educational area in this respect is huge.

The Nature of Contingency Measures

It will be recalled that contingency factors are those factors that interact with treatment conditions in bringing about educational outcomes. As with treatment conditions there are also three general classes of contingency factors: 1) learner characteristics, e.g., age, intelligence, sex, "learning styles", background of experience, 2) teacher characteristics, e.g., background of knowledge, background of experience, personality orientations, interests, and 3) setting characteristics, e.g., number of children in a class, size of classroom in relation to number of children in it, physical facilities, time of day. Historically, a great deal of effort has been directed to the measure-

ment of these factors, particularly the characteristics of the learner, for these are the factors which determine whether an instructional system will be effective or ineffective.

While many contingency factors involve direct measurement, e.g., age, sex, background of experience, number of children in the classroom, physical facilities, etc., many do not. Many contingency measures are indirect measures in that they are measures of qualities that are only inferred (constructs). Intelligence, learning style, anxiety, interest, and aptitude are concepts of this kind. One cannot see, hear, touch, feel, smell or taste these qualities; they can only be inferred from that which is observed. In this sense, the major difference between direct and indirect measurement is the reliance of indirect measurement upon that which is being observed as an indicator of the construct that is being measured and the reliance of direct measurement upon that which is being observed as a descriptor of that which is being measured. Since the issues involved in direct measurement have already been discussed the remainder of the present section is devoted to indirect measurement.

With such a basic difference existing between direct and indirect measures one would expect the measurement processes involved to be quite different. This is not the case. The steps or operations involved in direct and indirect measurement are exactly the same. So are the procedures involved in their construction. After identifying that which one wishes to measure, an appropriate measurement methodology is selected, situations to be sampled are identified, items are built and the measure is then tried out and evaluated. The issues of relevance, representativeness, fidelity and reliability are as crucial in indirect measurement as they are in direct measurement. The one difference that exists in these two classes of measures is the requirement that indirect measures must have some empirical evidence that they are in fact measuring that which they are supposed to be measuring. This calls for an attribute above and beyond those of relevance, representativeness, fidelity and reliability, namely, evidence of construct validity.

As with the concepts of relevance, representativeness and fidelity, construct validity also relates to the issue of whether a measure is measuring that which it is supposed to measure (constant error). They differ, however, in the kind of

evidence that is permissible in their support: the attributes of relevance representativeness and fidelity involve essentially non-empirical, analytic, judgemental evidence (the notions of "face" or "content" validity) whereas the attribute of construct validity requires experimentally obtained empirical evidence. This requirement makes validity evidence costly and difficult to obtain (for it requires a full-scale research program to do so), and as a consequence only the better "standardized" measures are likely to have it. Rarely will teachers be able to obtain validity evidence for the tests which they develop within a class.

Two kinds of empirical evidence are desirable in support of the construct validity of a measure: 1) correlation with another measure that is known to be a valid measure of the property under consideration (concurrent validity), and 2) the experimental verification of hypotheses which involve the concept being measured and which have used the test being considered in the research that has provided that verification (theoretical validity). Either or both kinds of evidence provides confidence that the measure "is in fact measuring that which it is supposed to be measuring."

One further comment about construct validity: just as the evaluation of an instructional system permits the clarification and testing of the conceptual framework which underlies it (the hierarchy of enabling objectives) the pursuit of construct validity permits the clarification and testing of the constructs used in a discipline. In this sense, obtaining evidence of construct validity is as much a conceptual or theory developing activity as it is a measurement activity. While most practicing educators will not be involved in the development of instruments to test concepts or theory, they need to be familiar with the idea of construct validity for whenever they use indirect measures in their research they will have to support them with evidence of this kind.

A Note on Prediction

Much of science is involved with prediction, for once prediction is possible it enables control. Educators are also concerned with prediction: who will pass and who will fail? Who is a good college risk and who isn't? Who is likely to benefit from curriculum A and who from curriculum B? To permit prediction there must be measurement, for one predicts whatever he is predicting from some set of measures. These may be as general as age or sex or socio-economic status, or as specific as number of

questions asked or number of comments made in a class. Of significance to the present discussion, however, is the fact that whenever any measure is included in a generally used prediction scheme there must be evidence of its predictive validity. Predictive validity, like construct validity, requires empirical evidence, i.e., that the measure does in fact predict that which it is supposed to predict.

A Note on the Selection of Available Measures for Use in Instructional Systems Evaluation and Research

Those who wish to consider using existing, standardized measures for purposes of evaluation will find the information in O. K. Buros's Mental Measurements Yearbooks (cf. 1959, 1965) of great value. A comprehensive coverage of all published tests in nearly all fields of education is available through this source, along with excellent reviews of the specific strengths and weaknesses of each. Reliability and validity data are included in the reviews.

Along with the Yearbooks Buros has published a 479-page index to Tests in Print (1961). With this reference source it is possible to locate rather easily the yearbook in which a particular test has been reviewed. To provide some notion of the range of subject areas covered by Buros the classification of tests included in the 1959 Yearbook is included as Appendix D.

Bibliography

Ahmann, J. S. and Grock, M. D. Evaluating Pupil Growth. Allyn Bacon, 1963.

Bellack, A. A., in collaboration with Hyman, R. T., Smith, Frank L., Jr., and Kliebard, H. M. The Language of the Classroom: Meanings Communicated in High School Teaching. Part II. U. S. Office of Education, Cooperative Research Project No. 2023. New York: Institute of Psychological Research, Teacher's College, Columbia University, 1965.

Baker, Eva. Construction of Prototype Items. Southwest Regional Laboratory for Educational Research and Development, 1967. (Mimeographed)

Buros, O. K. (Ed.). The Fifth Mental Measurements Yearbook. Highland Park, N. J.: Gryphon Press, 1959, 1965.

Buros, O. K. Tests in Print. Highland Park, N. J.: Gryphon Press, 1961.

Campbell, D. T. and Fiske, D. W. Convergent and discriminate validation by the multitrait-multimethod matrix. Psychology Bulletin. 1959, 56, 81-105.

Ebel, Robert L. Measuring Educational Achievement. Englewood Cliffs, N. J.: Prentice-Hall, 1965.

Festinger, L. and Katz, D. Research Methods in the Behavioral Sciences. N. Y.: Dryden, 1953.

G. ,e, N. L. (Ed.). Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

Gagné, R. The Conditions of Learning. N. Y.: Holt, Rinehart & Winston, 1964.

Gerlack, V. S., Schutz, R. E. and Baker, R. L. Measuring Educational Outcomes: Absolute vs. Relative Criteria. Southwest Regional Laboratory for Educational Research and Development, 1967. (Mimeographed)

Gornlund, N. E. Measurement and Evaluation in Teaching. N. Y.: MacMillan, 1965.

Guion, R. M. Personnel Testing. N. Y.: McGraw Hill, 1965.

Kerlinger, F. N. Foundations of Behavioral Research. N. Y.: Holt, Rinehart & Winston, 1965.

Lindquist, E. F. (Ed.). Educational Measurement. Washington, D. C.: American Council on Education, 1950.

Lindzey, G. (Ed.). Handbook of Social Psychology, Vol. 1. Cambridge, Mass.: Addison-Wesley, 1954.

Mussen, P. H. (Ed.). Handbook of Research Methods in Child Development. N. Y.: Wiley, 1960.

Sax, G. The Construction and Analysis of Educational and Psychological Tests: A Laboratory Manual. Madison, Wisconsin: College Printing and Typing Co., 1962.

Schalock, H. D. Issues in the conceptualization and measurement of teaching behavior. Paper presented as part of a symposium on the Prediction of Teaching Behavior, American Education Research Association, New York, February, 1967. (Mimeographed)

Standards for Educational and Psychological Tests and Manuals. Prepared by a joint committee of the American Psychological Assoc., the American Educational Research Association, and the National Council on Measurement in Education. Washington, D. C.: American Psychological Assoc., Inc., 1966.

Stanley, C. Measurement in Today's Schools. (4th Ed.) Englewood Cliffs, N. J.: Prentice-Hall, 1964.

Taba, Hilda, Levine, S. and Elzey, F. F. Thinking in Elementary School Children. U. S. Office of Education, Cooperative Research Project No. 1574. San Francisco State College, 1964.

Webb, E. J., Campbell, D. T. Schwartz, R. D. and Seechest, L. Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago: Rand McNally, 1966.

APPENDIX A

THE NATURE OF MEASUREMENT IN THE BEHAVIORAL SCIENCES

The Concept of Measurement

Lorge (1950, p. 533) has reported that in a sample of 2½ million words the term "measure" occurred more than 400 times and was used in 40 different ways. It was used to indicate the act of weighing, the balance in which the weighing was done, the grams that were used to balance, and the numeral that expressed the result. The term also referred to less exact instruments, processes, and units. In fact, any instrument that is used as a basis for comparison, even when that comparison involves the process of estimation or judgment generally carries the term "measure". Thus, as used popularly, measure not only refers to procedures that have precision, but also to acts of objective estimation, such as the estimation of beauty or the estimation of an individual's intelligence.

A Definition of Measurement

In its broadest technical sense, measurement is the assignment of numerals to objects or events according to rules (Stevens, 1951). As measurement is conceived within the physical sciences this is rather a straightforward definition: by using a device such as a ruler or a scale in accordance with relatively simple principles one can assign inches to a table top or ounces to a cup of flour without difficulty and with considerable (though never exact) accuracy. It can be a different matter in the behavioral sciences.

> Suppose that we ask a male judge to stand seven feet away from an attractive young woman. The judge is asked to look at the young woman and then to estimate the degree to which she possesses five attributes: niceness, strength of character, personality, musical ability, and intelligence. The estimate is to be given numerically. In the number system a scale of numbers from 1 through 5 is used: 1 indicating a very small amount of the characteristic in question and 5 indicating a great deal of the characteristic. In other words, the judge, just by looking at the young woman, is to assess how "nice" she is, how "strong" her character is, and so on, using the numbers 1,2,3,4, and 5 to indicate the amount of each characteristic she possesses.
>
> After the judge is finished, another male judge is asked to repeat the process with the same young woman. The numbers of the second judge are checked against those of the first judge. Then both judges similarly judge a number of other young women. (Kerlinger, 1965, p. 41)

The author goes on to point out that while this example may seem ridiculous as an illustration of measurement, it does in fact meet the definition of measurement. The judges assigned numerals to objects according to rules: the objects, the numerals and the rules for the assignment of the numerals to the objects were all specified, (the numerals were 1, 2, 3, 4, and 5; the objects were the young women) and the rules for the assignment of the numerals to the objects were contained in the instructions to the judges. What makes the example seem far fetched as an illustration of measurement is the fact that the properties of the characteristics being judged were not at all specified i.e., what is meant by "nice", "strong", "intelligent", etc., the rules for assigning the numerals to the properties were not clear, i.e. what properties had a value of 1, 2, 3, etc., and the conditions under which the observations were to be made were not spelled out. Without these minimal considerations "reproducebility of observations" is not likely to be reached. This of course is one of the first criteria of measurement that scientists and users of measures demand for without reproducibility a measure is of little value. To become reproducible, the properties of that which is to be observed, the rules by which numerals are to be assigned to those properties and the conditions under which the observations are to occur must be made explicit and public. Unfortunately, a definition of measurement contains no statement about the quality of the procedures involved in it.

While space will not permit an extended discussion of the intricacies involved in specifying properties, rules, context, et[illegible] e essential nature of each of these processes will be noted.

Specifying that Which is to be Measured

Perhaps the most basic of all concepts in measurement is the simple notion that in order to measure something one must know what it is he wishes to measure. Unfortunately, like the definition of measurement, this also is a deceivingly simple statement. Two requirements are inherent in it: 1) some concept of that which is to be measured exists and 2) some notion of the measureable properties of the concept exist. Put in another way, that which is to observed is dependent upon man's ability to conceive of it and, then, of his ability to observe it (Lorge, 1950, p. 536).

The development of concepts. An essential feature of any science is the development, extension and refinement of its concepts, for it is through its concepts that it gains its power. Conceptualization is a constant, cyclical data-dependent process wherein new data (observations) give rise to new concepts and new concepts give rise to new data. The cycle doesn't begin at a particular place, nor end; it is ever present and so long as man inquires it will forever be present.

The point that is critical here is that concepts "don't just happen" or "don't first exist." They are man made and they are constantly evolving.

For ages, man disregarded certain objects or their effects because he did not know of them or their behavior. He did not notice ultraviolet radiation, nor the fact that quartz reacted differently from glass to ultraviolet light. Nor did the notice electrical currents in the brain, or "unconscious" motivation or that some people can taste certain things and others can't. We now have these concepts, and they have opened an appreciable store of knowledge. We don't have them in the same form as they were conceived initially however. Atomic structure, circa 1967, is not atomic structure circa 1937. Neither is "intelligence" nor "motivating" nor "learning". Concepts come and go, maybe powerful or weak, but always they change. Moreover, whatever their form, they dictate that which is to be observed and measured.

The Identification of the Properties of Concepts. Having invented a concept, or having revised one, is not enough in itself to permit its measurement. The properties or characteristics of the concept must also be specified, for it is these that are in fact actually measured, i.e. receive the assignment of numerals to indicate quantity. Thus it is not sufficient simply to invent a concept such as intelligence, or achievement, or anxiety, or poverty, or cultural deprivation. One must also specify what it is that constitutes these concepts.

This constitutes one of the most difficult problems in measurement for as yet there are no rules governing how it is to be done, and no way of knowing whether it ever gets done. The procedure followed in specifying the properties of a concept has been labeled generally as "constitutive definition", (Margenau, 1950) (Torgerson, 1958), and refers to a procedure whereby a concept is simply defined with progressively lower-order concepts until one either encounters a set of concepts whose properties have been defined in other terms or reaches a point beyond which defining concepts can't be found. Thus, one may begin with a concept such as teacher behavior, break it down into its major components, e.g., caretaking, teaching, and routine-administration behavior, then break each of these down into their major components, etc. Ultimately a point is reached beyond which lower-order concepts are not applicable, and at that point the properties of a concept will have been specified so far as available knowledge permits. The assumption underlying this process is that it is at this level that measurement operations, i.e., the assignment of numerals to properties by rules, should take place. This whole process will be recognized of course as a procedure comparable to the hierarchical analysis of educational objectives that Dr. Twelker has reviewed in the institute.

One aspect of measurement that is bothersome to people is highlighted in the previous discussion, namely, the fact that any single measure can attend to only one or a few of many properties of a complex concept. No single measure of a teacher's behavior can ever

measure all of the properties or dimensions of her behavior; nor can a single measure of intelligence ever measure all of the rich and diverse properties of human intelligence, or a single measure of creativity ever measure all of the characteristics of human creativity. Even more bothersome is the fact that in science generally an effort is made to measure a property or characteristic of an object with little or no regard for any of its other properties or characteristics. The length of a table may be estimated without reference to its color, width, height, wood, style, or shape. Psychometrists try to make estimates of the "intelligence" of an adult with little or no consideration of his race, personality, or economic circumstance. Be this as it may, a single measure can at best give only an approximation to a concept, and if the concept is one which is complex, i.e. involves many properties, it will require many measures to yield an adequate approximation of it.

Specifying the Rules by Which Numerals are Assigned to that Which is to be Measured.

Assuming that the properties or characteristics of a concept have been identified (defined, conceptualized) there is still the task of assigning numerals to these characteristics in such a way as to indicate their quantity. This procedure must be governed by "rules" or "operations" which, like the properties of the concepts being measured, are made explicit and public. Only in this way can a measure be "reproducable", and thus admissible as objective information.

In thinking about the assignment of numerals to properties it needs to be recognized that a <u>numeral</u> is a symbol of the form: 1, 2, 3, ..., or I, II, III . . . . It has no quantitative meaning unless we give it such a meaning. It is simply a symbol of a special kind. The term numeral is used because measurement ordinarily uses numerals which, after being assigned quantitative meaning, become <u>numbers</u>. A <u>number</u> is a numeral that has been assigned quantitative meaning.

As used here the term "assigned" refers to the mapping of the objects of one set (numerals) to the objects of another set. In behavioral science research the members of one set are usually individuals, or properties of concepts which relate to individuals, and the members of the other set are usually numerals.

Rules govern the assignment of the objects of one set (numerals) to the objects of another set (persons or properties). In the behavioral sciences a "simple minded" rule might say: "Assign the numerals 1 through 5 to individuals according to how nice they are. If an individual is not at all nice, let the number 1 be assigned. Assign to individuals between these limits numbers between the limits." Another rule might be: "If an individual is male, assign him 1. If an individual is female assign her 0." Both rules of course assume

previous definition of concepts and isomorphism with reality." (Kerlinger, 1965, p. 413-417).

As with anything else the rules of measurement may be "good" or "bad". To the extent that they are good or bad, measurement is likely to follow suit. Many things are relatively easy to measure because the rules are easy to draw up and follow. To measure sex, for example, is easy since several simple and fairly clear criteria can be used to determine sex and to tell the investigator when to assign 1 and when to assign 0. It is also easy to measure certain other human characteristics: hair color, eye color, height, weight. Unfortunately, most human characteristics are much more difficult to measure, mainly because it is difficult to specify clearly the properties of the characteristic to be measured and to devise clear rules that govern the assignment of numerals to them. Nevertheless, rules of assignment must always govern the measurement process, and more and more attention must be directed to the specification of such if the behavioral sciences are to advance.

Specifying the Conditions Under which Measurement is to Occur

In scientific observations, whether direct or indirect, the conditions for observation must be carefully specified in terms of time, place, and circumstance. In physics and chemistry, observations at sea level at 25° centigrade may differ markedly from observations of the same thing at 0° centigrade and in an airplane 35,000 feet above sea level. In psychology, the behavior of an individual at 2 A.M. at his desk in his own home may differ markedly from his behavior at 10 A.M. at his desk in his office, or the behavior of a teacher may differ markedly in one subject area as compared to another or when working in a reading group with "good" readers as compared to "poor".

To control for this source of variance all measures necessarily must specify the conditions under which observations are made.

Measurement As a Sense Datum

As Lorge has pointed out (1950, p. 538-539), empirical measurement ultimately depends upon the occurrence of a sense datum and its interpretation by an observer. Weight, even though measured by machine, is ultimately related to the kinesthetic sensation of heavier and lighter. The machine--that is, scales--merely allows for a simple and relatively objective perception of the effects of weight. The geiger counter is a machine that enables the observer to perceive a specified class of effects by extending the range of human sensation. A test of intelligence serves the same purpose by standardizing a set of stimuli and a set of rules for applying the stimuli to the assessment of given properties of an individual. Other machines facilitate measurement by the control of systematic, chance, or erratic conditions, or by magnifying effects.

In this sense machines and tests simply allow more precise determinations of a particular class of effects. In the absence of instruments for extension of the senses, or for the control of conditions, human's observations are liable to error. Instruments are a means for approximating more closely the property under observation.

Unfortunately, there is a limit to the closeness of approximation in the measurement of a property by an observer or by an instrument. In calorimetry it is well known that the temperature of the measuring device affects the temperature of the material under observation, and in this respect observation must be corrected by calculation for the influence of the instrument. In intelligence testing an examiner can affect the score of the candidate by providing encouragement or its lack. In the physical sciences a great deal of attention is devoted to the reduction of the interaction of the instrument with the characteristic under observation but in the behavioral sciences this is rarely considered. Ultimately, however, the behavioral sciences also must learn to control for measurement effect. Until that time all one can do is to be aware that it exists and do all that is possible to reduce it.

Characteristics of Measurement in the Behavioral Sciences

In the previous section attention was given the fact that in any measurement the act of measuring is directed to the <u>properties of objects or concepts</u> rather than the object or concept itself. While this is true, it does not mean that the properties themselves are actually measured, for the properties of an object or concept are themselves concepts. What actually gets measured in the act of measuring are <u>indicants</u> of the properties of objects. This is the case in all measurement, whether in the physical or the behavioral sciences.

The problem of inference

Objects, processes, etc., or <u>concepts</u> of objects or processes, can be thought of as falling along a continuum of concreteness-abstractness. If one conceives of that which is to be measured in this way it becomes clear that some of the properties of objects, e.g., width, height, weight and hardness fall on the "concrete" end of the continuum and are therefore amenable to rather "direct" observation. This is not the case for such concepts as intelligence, morale, anxiety, hostility, or creativity. These are concepts that exist only by inference; they are assumed or inferred to exist because of certain observed regularities in behavior. As one moves away from the measurement of concrete, "natural" objects, that is, objects that are available to <u>direct</u> observation, and into the measurement of the properties of "inferred" objects or concepts, reliance upon indicants becomes heavier and heavier. As this happens uneasiness about measurement increases. As soon as relatively simple physical properties are left behind for more complex and elusive properties,

direct observation of properties is impossible. Hostility cannot be observed directly, nor can morale, anxiety, intelligence, creativeness, talent, etc. Such concepts require indirect measurement in that their properties or characteristics must always be inferred from observation of presumed indicants of those properties.

When using indirect measures indicants are taken as reflections of an underlying property or characteristic. They are seen as "something which points to something else." If a boy continually strikes other boys, we may say that his behavior is an indicant of underlying hostility. If someone's hands sweat excessively, we may say that he is anxious. If a child answers a certain number of items in an intelligence test correctly, we say he has a certain level of intelligence. In each of these cases, some identifiable behavior is assumed to be an indicant of an underlying property.

It is rather understandable that people view measurement as being rather shaky when it involves making inferences from observed behavior rather than directly observing properties like skin color, size or sex. To measure a child's cooperativeness, dependency, or imaginativeness is very different from measuring his height, weight, or wristbone development. The fundamental process of measurement is the same but the rules are much more difficult to prescribe. Inference dominates, and this creates one of the more vexing problems of psychological and educational measurement.

Operational Definitions

The use of operational definitions represent an attempt to overcome the problem of inference in the measurement of abstract concepts or constructs.[2] In essence, an operational definition is a definition that assigns meaning to a construct in terms of the operations that are used in measuring the indicants of that construct. Thus, if only one measure of a construct is used, intelligence may be defined as that which is measured by test X, or anxiety defined as the number of facial ties emitted over a given period of time and under specified conditions. As indicated earlier, a single measure rarely taps all of the properties of an object or a construct, and, as a consequence, more than one measure per construct usually should be obtained.

While operational definitions are indispensible ingredients in all scientific measurements they are dependent upon the conceptual

[2]Constructs are concepts which have the added meaning of being deliberately or consciously created for a special scientific purpose. Intelligence is a case in point.

activity that breaks out the properties of the concepts or constructs that are to be measured. Either activity by itself is relatively meaningless. In combination, however, they constitute the essence of the scientific enterprise, namely, the shuttling back and forth between the level of construct-hypothesis-theory building and systematic observation.

The Problem of Error Variance

Even though a measuring instrument reflects a careful conceptual effort, sensitive operational definitions, and a detailed statement of the rules for assigning numerals to properties, error will still enter the measures taken with it. Two people using the same yardstick to measure the same table will come up with different results. A chemist uses chemical balances so sensitive that they must be kept in another room to guard against the influence of body heat or of air currents set in motion by the chemist's movements; nevertheless, he will weigh material several times and still settle for an average as the "true" weight. The control and/or elimination of error in measurement is one of the most critical tasks of the scientist.

With physical measurements subject to error, it should not be surprising that behavioral science measurements are still more so. For example, attempting to measure intelligence with paper-and-pencil tests would seem to invite all kinds of error. And it does, but measures of intelligence also reflect to some degree the trait being measured; they are not totally inaccurate. If a person scores high on an arithmetic test, for example, it is reasonable to assume, despite the likelihood of error, that he is pretty good at arithmetic.

Essentially, this is the basic assumption of measurement in the behavioral sciences; any measure contains an element of error and an element of truth. Mathematically the assumption is that any obtained measure X is the algebraic sum of a true measure t and a measurement error e, or

$$X = t + e$$

Unfortunately, the statement is not entirely satisfactory. Logically, a "true" score is one that is not contaminated by any kind of error, but two kinds of error always exist in every measure: constant error and random error. By definition, a constant error is one that appears consistently in repeated measurement; random error is error that influences different measurements to different degrees. Now, if only random errors needed to be considered, the mean of repeated measurement would tend to give a good approximation of a "true" score because such errors are not correlated. But since constant error also occurs with each measurement taken, it influences the mean as much as it influences each individual measure. With a distinction between systematic, repeatable errors and those that vary randomly, it is possible to

rephrase the basic equation of testing as

$$X = s + e$$

with s (systematic measure) representing a composite of a true measure and any constant error. In this revision of the equation, e represents only the residual error which is rand. . and unpredictable. (Guion, 1966. p. 29).

In considering constant error, it needs to be recognized that two kinds of constant error exist. One can be attributed to the measuring instrument or to a measurement situation. In this case all objects or people measured are affected in the same way; that is, each measure in a total distribution is in error to the same degree. Other constant errors can be attributed to characteristics of an individual being measured. Repeated measure of a trait on one person may be consistently wrong for him to the same degree and in the same direction, for example, a person with chronic anxiety may consistently score 20 points too low on a given test every time he takes it. This is a constant error for him, but it does not influence other scores in the distribution.

This distinction between error that is constant within a whole array of measures and error that is constant for a specific individual is important, of course, only in relation to a whole set of measures. In the first case, the error is not likely to make a major difference in the interpretation of ones data since it does not upset relative positions within the distribution. In the second case the error has considerable practical significance because it does produce relative changes in the distribution. Statistically, this means that the first kind of error produces no variance, while the second kind of error has its own distribution and offers a variance of true scores.

Categorization and Scale Placement vs. Measurement

Ordinarily when one uses the term measurement he has as a referent something as manageable and stable as inches or yards or pounds or cents; something that can be added, subtracted, divided and multiplied. Unfortunately, measures in the behavioral sciences do not contain the characteristics which permit all of these operations, and because they don't there is serious debate about what one can in fact do (statistically and otherwise) with them. The purpose of the present discussion is to review briefly the concept of levels of measurement, identify the level of measurement reflected in most behavioral science measures, and then speak briefly to the implications of this for statistical analyses. In the discussion which follows I have drawn heavily from the work of Coombs (1953, pp 472-485) and Kerlinger (1966, pp 419-428).

Postulates of Measurement. Before discussing levels of measurement per se, some idea of three of the basic postulates of measurement is needed. This may sound a bit forbidding to those who are not attuned to mathematics, but in fact it is as straightforward as it is basic. A postulate is an assumption that is an essential prerequisite to carrying out some operation or line of thinking. In this case, it is an assumption about the relations between objects being measured. The three postulate may be written as follows:

1. Either $(a = b)$ or $(a \neq b)$, but not both

2. If $(a = b)$ and $(b = c)$ then $(a = c)$

3. If $(a > b)$ and $(b > c)$ then $(a > c)$

The first postulate says: "a is either equal to b or not equal to b, but not both." This postulate is necessary for classification. One must be able to assert either that one object is the same in a characteristic as another or that it is not the same. In measurement "the same" does not necessarily mean complete identity: it can mean "sufficiently the same to be classed as members of the same set." Saying that two boys are the "same" in maleness is in one sense accurate (they are both males) but it is likely that one boy may actually be more masculine than the other. This is a criterion matter. To be able to say that "the two are the same," one must only meet the criterion or the set of criteria that have been established as a measure of sameness. All criteria have one requirement, however: they must be sufficiently unambiguous to make classification possible, that is, to satisfy the condition the postulate states.

The second postulate says "If a equals b, and b equals c, then a equals c." Put in other terms, if one member of a universe is the same as another member, and this second member is the same as a third member, then the first member is the same as the third member. This postulate enables a researcher to establish the equality of set members on a particular characteristic.

The third postulate is of more immediate and practical importance for most measurement considerations in the behavioral sciences. It says, "If a is greater than b, and b is greater than c, then a is greater than c." This is the so-called transitivity postulate. Other symbols or words can be substituted for "greater than" (">") and "less than" ("<") e.g., "is at a greater distance than," "is stronger than," "precedes," "dominates," and so on. Most measurement in psychology and education depends on this postulate, for a goal of most measurement is to be able to assert ordinal or rank-order statements like "a has more of a property than b; b has more of the property than c; therefore a has more of the property than c."

The preceding statements may seem obvious, and in physical measurements the postulate is often satisfied: if stick a is longer than stick b, and stick b is longer than stick c, then stick a must be longer than stick c. If student a has more items right on a test than student b, and student b has more right than student c, student a must have more right than student c. But take the relation dominance: a may dominate b and b may dominate c, but it is possible that a does not dominate c. A wife may dominate her husband, and the husband may dominate their child, but the child may dominate his mother. If an investigator is studying dominance relations among children, he cannot simply assume that the postulate is correct. He must demonstrate that it is correct.

Most authors assign four levels of measurement: nominal, ordinal, interval and ratio, though (1953)

Levels of Measurement

1. The nominal scale. Measurement in its simplest form consists of substituting symbols or names for real objects. When measurement consists only in this mapping of objects into symbols, the symbols constitute a nominal scale. Thus a system which classifies occupations into families or the symptoms of patients into psychiatric classifications represents a nominal scale. To tie the discussion of levels of measurement to the postulates of measurement, the relation of "equal to" or "not equal to" must hold between objects on a nominal scale. This means that any pair of objects must clearly belong to the same class or not belong to the same class. In addition, the relation of equality must be symmetric and transitive. By symmetry is meant that if the relation holds between a and b, it also holds between b and a; symbolically, if a = b, then b = a. By transitivity is meant that if a = b and b = c, then a = c.

This level of measurement is so primitive that it is not always recognized as measurement, but it is a necessary condition for all higher levels of measurement.

2. Ordinal Scale. Sometimes the objects in one class of a nominal scale are more than just different from the members of another class-they may bear some kind of a relationship to them. One such relationship is that the members of the one class are more of something than the members of the other class, and it is meaningful to say that the relation "greater than" (>) holds between the members of one class and the members of the other in relation to some property. When this relationship holds for all members of the two classes the result is an ordinal scale.

3. The Interval and the Ratio Scale. In the two scales discussed heretofore--nominal, and ordinal--the elements of the system were classes of objects, and the relationships were relationships of

equality and greater than. Nothing was said about a concept of distance between classes. Thus, although a may have been observed to be greater than b, and b greater than c, nothing has been said about a being greater than b by a larger amount than b was greater than c. In interval and ratio scales the concept of distance between classes enters the picture, and with it significant increase in the power of measurement. The increase in power derives from the fact that once information is available on how large the intervals are in the property of the object being measured it then becomes possible to apply arithmetic operations to them. Interval scales provide data that are subject to addition and subtraction, (though it should be noted that with this level of data it is not quantities or amounts that are added or subtracted, but only intervals or distances). It is only when one is able to use ratio scales, however, that all of the arithmetic functions can be applied, that is, multiplication and division as well as addition and subtraction. This is made possible by the fact that a ratio scale, in addition to possessing the characteristics of nominal, ordinal, and interval scales, has an absolute zero that has empirical meaning. If a measurement is zero on a ratio scale, there is a basis for saying that some measured object has none of the property being measured. Numbers on the scale indicate the actual amounts of the property being measured; if a ratio scale of achievement existed, for example, it would be possible to say that a pupil with a scale score of 8 had an achievement twice as great as a pupil with a scale score of 4. For this reason ratio measurement is the ideal of all scientists.

Unfortunately, measures in the behavioral sciences are far from reaching the ideal of ratio measurement, and there is no reason to believe that they even will. Most measurement is of the nominal and ordinal variety, though many psychological and educational measures are treated as if they approximated interval measurement. The implications of this are two fold: 1) behavioral science measures of necessity each in precision, and 2) the statistical analyses that can be applied most to behavioral science data are limited. Dr. Beaird speaks to this point at length in the chapters upon analysis.

SUGGESTED READINGS
ON
THE NATURE OF MEASUREMENT IN THE BEHAVIORAL SCIENCES

1. For the mathematical foundations of measurement see Stevens, SS. Mathematics, measurement and psychophysics. In S.S. Stevens (Ed.) Handbook of Experimental Psychology. N.Y.: Wiley, 1951. Pp 1-49.

2. For good introductory references to the concept of measurement see:

   Kerlinger, F.N. Foundations of Behavioral Research. N.Y.: Holt, Rinehart and Winston, 1965. Chapters 1, 3, and 23.

   Lorge, I. The fundamental nature of measurement. In E.F. Lindquist (Ed.) Educational Measurement. Washington D.C.: American Council on Education, 1950. Chapter 14.

   Peak, Helen. Problems of objective observation. In L. Festinger and D. Katz (Eds.) Research Methods in The Behavioral Sciences. N.Y.: Dryden, 1953. Chapter 6.

   Coombs, C.H. Theory and methods of social measurement. In L. Festinger and D. Katz (Eds.) Research Methods In The Behavioral Sciences. N.Y.: Dryden, 1953 Chapter 11.

3. For works interested in measurement from the point of view of the philosophy of science, see Campbell N.R. Foundations of Science: The Philosophy of Theory and Experiment. N.Y.: Dover, 1957.

Ellis, B. Basic Concepts of Measurement. Cambridge, England: Cambridge Univ. Press, 1966.

Margenau, H. The Nature of Physical Reality. N.Y.: McGraw-Hill, 1950.

Torgerson, W. Theory and Methods in Scaling. N.Y.: Wiley, 1958.

Rose, S. Logical Foundations of Psychological Measurement. A Study in the Philosophy of Science. Copenhagen: Scandinavian Univ. Press, 1964.

APPENDIX B

CLASSES OF MEASURES USED IN THE BEHAVIOR SCIENCES, THE NATURE OF THE DATA THAT DERIVE FROM THEM, AND SOME COMMENTS AS TO THE ADVANTAGES AND DISADVANTAGES OF EACH

Class of Measure	Data Form	Data Level	Particular Strengths	Particular Weaknesses
OBTRUSIVE MEASURES				
Interviews				
Structured				
- using fixed alternative items (agree-disagree, forced choice, rank order)	Category frequency counts e.g., total agreements or total "don't knows"	Nominal	Constant stimulus conditions; opportunity to clarify misunderstanding; ease of data preparation	Lacks the flexibility that is possible in an interview; data costly to obtain
	Total scores, e.g., number right or wrong	Ordinal		
	Rankings, e.g., preferences	Nominal Ordinal		
- using open ended items	Category frequency counts (through content analysis)	Nominal	Constant stimulus conditions; opportunity to clarify misunderstanding	Data costly to obtain and prepare for analysis; requires extensive coder training
	Ratings (through content analysis)	Nominal Ordinal		
Unstructured	Category frequency counts (through content analysis)	Nominal	Freedom to pursue a topic as the situation dictates	Stimulus only generally known, thus comparability of data is questionable; data costly to obtain and prepare; extensive coder training required
	Ratings (through content analysis)	Nominal Ordinal		

Class of Measure	Data Form	Data Level	Particular Strengths	Particular Weaknesses
<u>Systematic Observation</u> (under either "naturalistic" or "experimental" conditions)				
Diary Records	Category frequency counts (through content analysis)	Nominal	None, except so far as as it is better than <u>no</u> recording of an observation	Subject to the error of recall; subject to limitations of longhand recording; observations usually not focused
	Ratings (through content analysis)	Nominal, Ordinal		
Check Lists	Category frequency counts	Nominal	Provides focus and/or order to one's observations	Cumbersome for recording more than a few items; capable of handling a relatively limited range of data at one time
Rating Scales	Ratings	Nominal, Ordinal	Provides focus and/or order to one's observations; permits the summarization of large amounts of information in one score	Uncertainty of what the rating means, i.e., what cues/criteria it is based upon, thus comparability questionable; generally unreliable and lacking in evidence of validity
Running Records Using Preconceived Category Sets	Category frequency counts; (sequential ordering of behavior)	Nominal in sequence	Provides focus to observations; permits the handling of a large amount of information at one time; permits the sequential ordering of the behavior observed	Requires extensive observer training; data costly to obtain

APPENDIX B Continued

Class of Measure	Data Form	Data Level	Particular Strengths	Particular Weaknesses
Standardized Objective Measures				
Intelligence and Aptitude Measures	Part test and/or total test scores	Ordinal Interval	Constant Stimulus conditions; normative data available for reference; reliability and validity evidence available	Culturally and often symbolically dependent; limited sampling of situations requiring the application of "intelligence"
Achievement Measures	Part test and/or total test scores	Ordinal Interval	Constant Stimulus conditions; normative data available for reference; reliability and validity evidence available	Culturally and often symbolically dependent; limited sampling of situations requiring the application of "intelligence" which reflect the level of achievement
Personality, Attitude, Value and Interest Measures	Part test and/or total test scores	Ordinal	Constant Stimulus conditions; normative data available for reference; reliability and validity evidence available	Dependent upon the assumption that what one says is related to what one does or is, validity data limited

APPENDIX B Continued

Class of Measure	Data Form	Data Level	Particular Strengths	Particular Weaknesses
Standard Projective Measures				
Association Completion	Part test and/or total test scores Category frequency counts; ratings	Ordinal Nominal Ordinal	Standardized Stimuli that invite free response; some involve "realistic" stimuli usually invite strong personal involvement	Analysis is highly dependent upon idiosyncratic interpretations; little evidence of reliability or validity
Teacher Made Tests				
Short answer tests (using agreement - disagreement, forced - choice a rank - order stereotypes)	Part test or total test scores	Ordinal	Constant Stimulus conditions; easy to administer; easy to score; opportunity to standardize	Difficult to develop; response mode limited; test level limited to knowledge or described behavior in a situation response test
Essay tests and written documents	Ratings	Nominal Ordinal	Offers opportunity to assess writing skills and integration; relatively easy to develop	Limited sampling of situations which relate to the property being tested; scoring difficult; little opportunity to standardize
Products (e.g., articles such as furniture, clothes, models of machines; experiments, etc.)	Ratings	Nominal Ordinal	Offers opportunity assess application skills; easy to develop	Limited sampling of situations which relate to the property being tested; scoring difficult; little opportunity to standardize

APPENDIX B Continued

Class of Measure	Data Form	Data Level	Particular Strengths	Particular Weaknesses
NONOBTRUSIVE/MEASURES				
Physical Traces e.g., wear on the floor in front of a display; wear on library books, accumulation of garbage or bear bottles, unspecified products (see above)	Any form of evidence as to erosion or accretion		Inconspicuous; uncontaminated by the measurement	Of limited utility in the behavioral sciences; gross; validity and reliability hard to establish
Documents and Products	Category frequency counts (through content analysis); ratings	Nominal Ordinal	Rich sources of information; relatively easy to obtain; uncontaminated by the measurement process	Limited sampling of situations which relate to property being tested; scoring difficult; little opportunity to standardize
Simple Observation	Anecdotal records; ratings	Nominal Ordinal	Some record of some things, which probably is better than nothing	Simply not enough structure nor clear enough specification of rules for assigning numerals to observations to permit it to be called measurement
Contrived Observation				
(Hidden Hardware)	Category frequency counts (through content analysis)	Nominal	Permanent record of behaviors thereby permitting leisurely and multiple analysis; uncontaminated by the measurement process	Violation of privacy; unethical
	Ratings (through content analysis)	Nominal Ordinal		

APPENDIX C

MULTIPLE CRITERION MEASURES FOR EVALUATION OF EDUCATIONAL OUTCOMES*

I, Indicators of Status or Change in Cognitive and Affective Behaviors of Student in Terms of Standardized Measures and Scales (Obtrusive Measures)

Standardized achievement and ability tests;

Standardized self inventories designed to yield measures of adjustment, appreciations, attitudes, interests, and temperament;

Standardized rating scales and check lists for judging the quality of products in visual arts, crafts, shop activities, penmanship, creative writing, exhibits for competive events, cooking, typing, letter writing, fashion design, and other activities;

Standardized tests of psychomotor skills and physical fitness.

II. Indicators of Status of Change in Cognitive and Affective Behaviors of Students by Informal or Semiformal Teachermade Instruments or Devices (Obtrusive Measures)

Incomplete sentence technique: categorization of types of responses, enumeration of their frequencies, or rating of their psychological appropriateness relative to specific criteria.

Interviews: frequencies and measurable levels of responses to formal and informal questions raised in a face-to-face interogation.

Peer nominations: frequencies of selection or of assignment to leadership roles for which the sociogram technique may be particularly suitable.

Questionnaires: frequencies of responses to items in an objective format and numbers of responses to categorized dimensions developed from the content analysis of responses to open-ended questions.

---

* From a paper by Metfessel, N.S. and Michael, W.B. A paradigm involving multiple criterion measures for the evaluation of the effectiveness of school programs. Paper read at the 1967 AERA conference in New York. (Mimeographed).

Self-concept perceptions: measures of current status and indices of congruence between real self and ideal self--often determined from use of the semantic differential or Q-sort techniques.

Self-evaluation measures: student's own reports on his perceived or desired level of achievement, on his perceptions of his personal and social adjustment, and on his future academic and vocational plans.

Teacher-devised projective devices such as casting characters in the class play, role playing, and picture interpretation based on an informal scoring model that usually embodies the determination of frequencies of the occurrence of specific behaviors, or ratings of their intensity or quality.

Teacher-made achievement tests (objective and essay), the scores on which allow inferences regarding the extent to which specific instructional objectives have been attained.

Teacher-made rating scales and check lists for observation of classroom behaviors: performance levels of speech, music, and art; manifestation of creative endeavors, personal and social adjustment, physical well being.

III. Indicators of Status or Change in Student Behaviors Other than Those Measured by Tests, Inventories, and Observation Scales in Relation to the Task of Evaluating Objectives of School Programs (Nonobtrusive Measures)

Absences: full-day, half-day, part day and other selective indices pertaining to frequency and duration of lack of attendance.

Anecdotal records: critical incidents noted including frequencies of behaviors judged to be highly undesirable or highly deserving of commendation.

Appointments: frequencies with which they are kept or broken.

Articles and stories: numbers and types published in school newspapers, magazines, journals, or proceedings of student organizations.

Assignments: numbers and types completed with some sort of quality rating or mark attached.

Attendance: frequency and duration when attendance is required or considered optional (as in club meetings, special events, or off-campus activities).

Autobiographical data: behaviors reported that could be classified and subsequently assigned judgmental values concerning their appropriateness relative to specific objectives concerned with human development.

Awards, citations, honors, and related indicators of distinctive or creative performance: frequency of occurrence or judgments of merit in terms of scaled values.

Books: numbers checked out of library, numbers renewed, numbers reported read when reading is required or when voluntary.

Case histories: critical incidents and other passages reflecting quantifiable categories of behavior.

Changes in program or in teacher as requested by student: frequency of occurrence.

Choices expressed or carried out: vocational, avocational, and educational (especially in relation to their judged appropriateness to known physical, intellectual, emotional, social aesthetic, interest, and other factors).

Citations: commendatory in both formal and informal media of communication such as in the newspaper, television, schoo assembly, classroom, bulletin board, or elsewhere (see Awards).

"Contacts": frequency or duration of direct or indirect communications between persons observed and others.

Disciplinary actions taken: frequency and type.

Dropouts: numbers of students leaving school before completion of program of studies.

Elected positions: numbers and types held in class, student body, or out-of-school social groups.

Extracurricular activities: frequency or duration of participation in observable behaviors amenable to classification such as taking part in athletic events, charity drives, cultural activities, and numerous service-related avocational endeavors.

Grade placement: the success or lack of success in being promoted or retained; number of times accelerated or skipped.

Grade point average: including numbers of recommended units of course work in academic as well as in non-college preparatory programs.

Grouping: frequency and/or duration of moves from one instructional group to another within a given class grade.

Homework assignments: punctuality of completion, quantifiable judgments of quality such as class marks.

Leisure activities: numbers and types of; times spent in; awards and prizes received in participation.

Library card: possessed or not possessed; renewed or not renewed.

Load: numbers of units or courses carried by students.

Peer group participation: frequency and duration of activity in what are judged to be socially acceptable and socially undesirable behaviors.

Performance: awards, citations received; extra credit assignments and associated points earned; numbers of books or other learning materials taken out of the library; products exhibited at competitive events.

Recommendations: numbers of and judged levels of favorableness.

Recidivism by students: incidents (presence or absence or frequency of occurrence) of a given student's returning to a probationary status, to a detention facility, or to observable behavior patterns judged to be socially undesirable (intoxicated state, dope addiction, hostile acts including arrests, sexual deviation).

Referrals: by teacher to counselor, psychologist, or administrator for disciplinary action, for special aid in overcoming learning difficulties, for behavior disorders, for health defects or for part-time employment activities.

Referrals: by student himself (presence, absence, or frequency).

Service points: numbers earned.

Skills: demonstration of new or increased competencies such as those found in physical education, crafts, homemaking, and the arts that are not measured in a highly valid fashion by available tests and scales.

Tardiness: frequency of.

Transiency: incidents of.

APPENDIX D

CLASSIFICATION OF TESTS IN THE
FIFTH MENTAL MEASUREMENTS YEARBOOK (1959)

APPENDIX E

CONSTRUCTION OF PROTOTYPE ITEMS

Eva Baker

Southwest Regional Laboratory for Educational Research and Development
11300 La Cienega Boulevard
Inglewood, California

CONSTRUCTION OF PROTOTYPE ITEMS

After objectives have been phrased in terms of the learner's behavioral response to stated stimulus or presentation conditions, even further clarification for the teacher or producer of instructional materials is requisite. The next essential step involves the production of a prototype item designed to measure the achievement of the objective. Provision of such a prototype performs a variety of functions. First, it may clarify any ambiguities in the stimulus or presentation conditions stated in the objective. Next it provides a model for additional items for the complete posttest, so that the test constructor has a definite guide and can generate items with more efficiency.

"Prototype" means that the item must be a specific example of the general type of behavior described by the objective. This in turn implies that a particular objective must describe classes of stimuli and responses clearly enough that members (i.e., items) of the class can be designed from this more general description.

The way to tell if this is in fact the case is to inspect the objective and item on three particular dimensions: <u>learner response</u>, <u>test directions</u>, and <u>presentation conditions</u>.

The factor of prime importance is the learner response. In many instances there is little difficulty incurred or subtlety required in making the judgment on the basis of <u>response</u>. If an objective calls for the selection of answers by circling correct definitions, the item must provide an opportunity for the learner to circle alternatives. Suppose one objective calls for the learner to use a particular strategy when solving in writing problems of a given type. If the test item only provides a place for the learner to record the solution to the problem but no place for him to indicate the strategy that he used, then the item would <u>not</u> be a prototype of the behavior in the objective. Because the objective required the use of a <u>strategy</u> as well as a solution, some way of recording the <u>process</u> used to solve the problem must also be supplied.

Here is a specific example of the necessity for objective-prototype correspondence. Does the item below satisfy the response called for by the objective? Circle yes or no below.

<u>Objective</u>: The learner will describe in an essay the way to make a cake.

<u>Test Item</u>: <u>Directions</u>: Circle the best answer.
The way to make a cake is to
1. beat eggs separately and fold in

2. combine dry ingredients before adding liquid
3. use a mix

Yes No

* * *

Clearly the response called for in the item does not conform to the specifications in the objective. An appropriate item might be:

IN A 500-WORD ESSAY, DESCRIBE THE STEP BY STEP PROCEDURE NECESSARY TO MAKE A GERMAN CHOCOLATE CAKE.

Here, the behavior called for is the same described in the objective. The actual response stipulated in the objective must be required by the item.

The second dimension, importance of test directions, can be well illustrated by considering this next objective.

<u>Objective</u>: Given three sets of words and a stimulus word, the student will underline words with the same initial sound as the stimulus word.

<u>Item</u>: <u>Directions</u>: Underline all words which sound like the first word.

BOY: BIRD TOY BAT

In this case, the directions are ambiguous and the learner could reasonably underline all three words if he considered rhyming sounds as well. In this case, the underlining looks like the appropriate response to the objective. The item might have appeared to be satisfactory had the directions for the learner not been included, but the same exact behavioral response (underlining words) would reflect a completely different process than that specifically described by the objective. Correct directions might be as follows:

<u>Directions</u>: Underline all words which have the same starting sound as the first word

A third focus in matching item to objective is observing if the <u>presentation conditions</u> specified in the objectives are actually those exemplified by the item. If the objective says the learner is to be given special information to assist his performance, the item should be checked to see if such necessities have been provided.

For example, one's objective may be that:

THE LEARNER WILL SELECT THE CORRECT ESTIMATE OF TIME, GIVEN THE SPEED AND DISTANCE OF A TRAIN ON A TRIP.

The item should be checked to see if necessary information is given.

For the next objective what presentation conditions must be met? Write your answer below.

THE LEARNER WILL DESCRIBE IN WRITING AREAS OF POPULATION DENSITY AFTER INSPECTING A MAP OF TOPOGRAPHY OF A GIVEN REGION.

____

____

*　　　*　　　*

The map that reflects the physical features of the designated area must be provided. Prototype items should be checked, then, in three ways, all of which are designed to determine if the item is in fact an actual member of the class of stimulus-response combination described by the objective.

List below the three dimensions one should consider when verifying objective-prototype item correspondence.

1. ________________________________
2. ________________________________
3. ________________________________

*　　　*　　　*

The three dimensions are:

1. The learner's actual response to the item should be compared to the response called for in the objective.
2. The directions given the learner to answer the item should reflect the intention of the objective.
3. The item should exhibit or provide for the presentation conditions given in the objective.

The use of this miniature system of verification should promote high objective-prototype correspondence and thus allow the instructional planner to proceed with more confidence.

For each of the following pairs of objectives and prototype items, verify that the item calls for behavior appropriate to the objective by inspecting the learner response desired, the test directions, and the presentation conditions. Circle "yes" below if the objective and item correspond in these three respects. Circle "no" if they do not correspond and check those dimensions which do not correspond.

Objective: THE LEARNER WILL LABEL A MAP WITH APPROPRIATE NATURAL RESOURCES FOR LATIN AMERICAN NATIONS.

Item: Directions: On a piece of paper list the natural resources of the following countries:
A. Chile
B. Paraguay
C. Guatemala

YES NO

___________Response

___________Directions

___________Presentation Conditions

* * *

You should have circled no for this pair and checked response and presentation conditions. The item makes no provision for the map upon which the students are to place the names thereby lacking presentation conditions. The response desired is also at variance with the objective since labeling a map requires a knowledge of what resources go with which region rather than a simple listing.

Consider this next pair. Answer yes or no below and check the dissonant dimensions, if any.

Objective: THE LEARNER WILL CIRCLE CORRECT EXAMPLES OF PATTERN 4 SENTENCES WHEN GIVEN A SET OF FOUR SENTENCES.

Item: Directions: Circle all examples of pattern four sentences.
A. The girl was pretty.
B. Quietly the wave broke on the shore.
C. The airplanes roared.
D. Children are often playful.

YES NO

___________Response

___________Directions

___________Presentation

* * *

You should have answered "Yes," because the objective and item do correspond, in terms of learner response, directions and stimulus conditions. Try this next set. Answer below.

Objective: THE STUDENT WILL CIRCLE DESCRIPTIONS OF PROBLEMS WHEN PRESENTED WITH BOTH PROBLEM AND NON-PROBLEM DESCRIPTIONS.

Item: Directions: Which are problems? Circle the letter in front of each problem below.
A. Susan has lost her purse.
B. Sam has an ice cream.

YES NO

___________Response

___________Directions

___________Presentation Conditions

* * *

You should have answered "Yes," because again the item and objective match on the three stated criteria. Does the following item adequately reflect the objective? Answer yes or no below and check those dimensions which do not correspond.

Objective: THE LEARNER WILL INFER THE POLAR QUALITY OF MAGNETISM AND CORRECTLY PLACE OPPOSITE POLES OF BAR MAGNETS TOGETHER.

Item: Directions: From the picture provided circle the two ends of the magnet which attract each other.

YES NO

___________Response

___________Directions

___________Presentation Conditions

* * *

You should have circled "No" in response to this objective - item combination, and checked all 3 dimensions since the presentation conditions (pictures rather than magnets), directions, and learner response (circling rather than manipulating) differ from what is stated in the objective.

For the next objective to write a prototype item including directions which would exemplify all necessary stipulations of the objective. Write your item below.

GIVEN FOUR SIMPLE 2-DIGIT ADDITION PROBLEMS, THE LEARNERS WILL WRITE THEIR ANSWERS.

Directions: ____

____

____

*   *   *

Your item had to include four 2-digit problems. Your directions should have sounded something like this:

Directions: Write the answer to each problem.
This would be sufficient if you provided addition (+) signs for each of the problems. Otherwise you should have stated something like this:

Directions: Add the numbers in each problem and write the sum on your paper.

The problems could look like this:

$$\begin{array}{r} 71 \\ +35 \\ \hline \end{array}$$

or like this:

71 + 35 =

or like this:

71 plus 35 is ______.

Yet any of these would be appropriate to the stated objective.

The writing of objectives and the subsequent clarity provided by producing prototype items reflect a procedure considered mandatory before any serious instruction is possible. Objectives should be stated in terms of the learners' behavior, to be exhibited under particular presentation conditions specified in the objective.

Items which purportedly measure the attainment of the objectives should be particular instances of the class of stimuli and responses described in the objective. The actual correspondence between objective and

item must be verified by using three dimensions: learner response, directions and presentation conditions.

After reading through this paper it is hoped that you will be able to use specific objectives on your instructional planning. Your objectives should state the specific learner response to given presentation conditions. When confronted with objectives which do not meet this standard, you should be able to modify them. Further, you should be able to identify and write test items which exemplify the objective's specifications. Such abilities will unquestionably result in a greater productivity and efficiency in your instructional planning.

APPENDIX F

MEASURING EDUCATIONAL OUTCOMES:

ABSOLUTE VS RELATIVE CRITERIA

Vernon S. Gerlach
Richard E. Schutz
Robert L. Baker

Southwest Regional Laboratory for Educational Research and Development
11300 La Cienega Boulevard
Inglewood, California

RELATIVE VS. ABSOLUTE CRITERIA

There are two basic methods of describing a learner's performance. The first, and perhaps most familiar to teachers, is that method by which a given learner is compared with other learners. The teacher who does this uses a relative criterion. The second is that method by which a learner's performance or achievement is compared with an arbitrary, pre-specified standard. This is the use of an absolute criterion.

The teacher who "grades on the curve" is using a relative criterion. He has no A Priori knowledge of how much the best, or the poorest, or the mean learner will achieve. He does know that, irrespective of the performance of any individual or of the group as a whole, a pre-specified number of pupils will receive an A and a pre-specified number will "fail." The cutting point for the A's or the failures is not known in advance. It is determined on the basis of the performance of all the pupils in a given group. Whenever performance or achievement is assessed in this manner, the teacher is using a relative criterion. This is, the criterion is how well the student performed relative to the group.

Most teachers have used percentiles at one time or another. This is another example of the use of a relative criterion. Consider a typical spelling test.

Prior to the administration of the test, no one knows whether or not any of the pupils will spell all the words correctly. It is conceivable that several pupils will score 100% correct, and it is equally conceivable that no one will; when one uses a relative criterion, such as a percentile rank, this factor is irrelevant. The assessment is concerned solely with ranking each pupil on the basis of his test score: this rank, in the case of percentiles, tells one what percent of all the pupils any one pupil surpassed.

To put it differently, a pupil with a percentile rank of 90 surpassed 90 of the 100 pupils in performance on the spelling test. However, there is no possible way of knowing how many words he spelled correctly, given only his percentile score or rank.

1. Why is it impossible to know a pupil's actual (or raw) score on a spelling test, given only his percentile score?

::: ::: :::

1. Percentile score is a relative criterion. It compares pupils. It does not tell how he has achieved with respect to a possible "perfect" score.

****

One cannot distinguish between relative and absolute measures merely on the basis of whether performance is reported in terms of a raw score. Raw scores may very easily be used in a relative criterion context (and very often they are used in this manner). Dr. Dale administers an algebra test consisting of 43 items. Last year the highest score was 39; and 28 was the cut-off point between passing and failing. Everyone who scored 37,38, or 39 was given an A. This year the high score was 42, the cut-off point was 32, and everyone who scored 41 or 42 received an A. Dale is concerned with how a given pupil compares with the other pupils in his class as far as performance on the 43-item test is concerned. He does not decide in advance how many of the 43-items a good algebra student must answer correctly.

He does not use an absolute criterion. He does use a norm-referenced criterion.

2. Think of a standardized achievement test, such as the Metropolitan or the Stanford. Is it a norm-referenced measure or not?

::: ::: :::

2. As ordinarily used in a classroom situation, a standardized achievement test is a norm-referenced measure.

****

But someone may ask, what about the fact that a sixth-grade teacher knows, before the test is administered, how many points a pupil must score in order to achieve a result of 7.0 (Grade Equivalent) on the arithmetic reasoning section? The answer is very simple. The test was standardized, i.e., norms were established for a large sample of pupils throughout the country--and the assumption is made that the pupils in this sixth grade class are similar to the pupils in the large sample. Before the test was standardized no one knew exactly where the cutting point would be for 7.0 or for any other grade equivalent. If, however, a teacher decided to pass only those who achieved a grade equivalent score of 7.0 or higher, he would be using this test in the absolute sense, since a grade equivalent of 7.0 is equal to a specific raw score on the form being used. A teacher uses an absolute criterion when he specifies in advance what the minimum acceptable performance is. Furthermore, the performance must be described in terms of the objectives, and not in terms of how the pupil compares with other pupils.

If a teacher wants the pupils to give the sums of the 100 addition facts, he uses an absolute criterion. He keeps teaching the pupils--every pupil! --until all the pupils in the class have attained the objective. The teacher knows before he administers any test just exactly what a "passing" performance will be: 100 out of 100.

Another teacher may be teaching pupils to capitalize correctly. He uses proof-reading type exercises, in which pupils must learn to identify

and correct errors in capitalization. Later, he administers a test containing items similar to the practice exercises. Pupils who achieve a perfect score on the test go on to learn other things, while those who do not achieve a perfect score receive additional instruction in capitalization. This teacher is using an absolute criterion. Similarly, if he had decided A Priori that any pupil who achieves a score of 90% correct or higher will have attained the objective, he would have been using an absolute criterion. However, if he had decided that the 90% of the pupils in the class who receive the highest scores will have attained the objective, he would have been using a relative criterion.

3. A teacher prepares a 100-item test in addition of common fractions. He decides that anyone who attains a score of 80 or more correct will have achieved the stated objective. Is this a relative or absolute norm?

::: ::: :::

3. This is an absolute. Each pupil is compared with an 80% or better performance. He is not compared with other pupils.

****

4. The teacher uses the test in 3, above. He decides that anyone who can solve 10 consecutive problems without an error will have attained the objective. One pupil may solve the first ten correctly; he is finished. Another may solve nine, miss one, solve seven, miss one, and so on, and not solve ten consecutive problems correctly until he reaches the 75th example. Another misses the first three items and then solves "ten-in-a-row" correctly. The teacher says, "All three of these pupils have attained the objective." Is the teacher using an absolute or a relative criterion?

::: ::: :::

4. Again, an absolute criterion. Pupils are compared only with a pre-specified standard; ten consecutive correct solutions. The pupils are not compared with each other.

****

5. "All who achieve a score of 90% on this test may move on to the next chapter." Relative or absolute?

::: ::: :::

5. Absolute.

****

6. The top 90% on this test will pass. The lowest 10% will have to take remedial work on this chapter after school. Relative or absolute?

::: ::: :::

6. This is a relative criterion. No one knows how high the passing score will be. Pupils, therefore, are being compared with each other, not with a pre-specified behavior.

****

7. "When 90% of the pupils achieve a score of 80% or higher, we will move on to the next unit." Relative or absolute?

::: ::: :::

7. Absolute, actually, there seems to be a relative standard here--90% of the pupils--but this is misleading. The critical element here is "80% or higher" on the test. This is a pre-specified, absolute requirement. The 90% of the pupils simply limits, or lowers, the absolute standard. The <u>entire class</u> will go to the next unit when a pre-specified standard is attained.

****

If, however, one said "The first 90% of the class to attain a score of 80% or higher will pass, while the others will not," one would be mixing two standards. While the 80% or higher remains absolute, a time element is introduced as an additional criterion. Pupils will be compared with each other with respect to the time required to attain a score of 80%. No one can tell, in advance, how long it will take the slowest of the top 90% to attain the objective. Nor does it make any difference whether it takes a week, a month, or a semester. As long as he is one of the first 90% to achieve the specified score, he has "made it." But if the pupil is one of the slowest 10%, he does not pass. Thus with two standards you have (1) an absolute standard with respect to score on a test, (2) modified by a relative standard with respect to time.

V - SYSTEMS DEVELOPMENT

Evaluation Sheet

We need to know how well the ideas and issues in this manual are communicated to you. You are the test audience for this material. To remove or strengthen the weak spots, to retain or improve the strong ones (if any), an account of your learning experience as you read these sections is crucial. Use this sheet as you study and jot down your reactions. Don't be concerned with typographical errors. We're concerned with how the message is coming through.

SECTION #________ TITLE__________________________

Give topic, paragraph or sentence, page	Importance Rate 1 2 3 4 5 Low - High	Clarity Rate 1 2 3 4 5 Low - High	Suggested Improvements

Over-all Rating of the Section: Use a 5-point Scale

Important				
1 Very Low	2	3	4	5 Very High

(Circle one)

Understandable				
1 Very Low	2	3	4	5 Very High

(Circle one)

PLEASE PUT ANY ADDITIONAL COMMENTS ON BACK

SECTION V

Instructional System Development

Table of Contents

Stages:

Cont'd Bibliography

Tables

SECTION V

Instructional System Development

Dale G. Hamreus

Thus far in the institute you have learned (1) how to state behavioral objectives that clearly indicate what terminal or outcome performances are expected of the learner following instruction; and (2) how to determine the specifications for achieving the terminal behaviors in (1). The next step to be presented in this section, is to develop the instructional materials and methods that are required to comply with (2) in fulfilling (1). This instructional materials development produces what is termed the instructional system.

In this institute a system is defined as any combination of persons, objects, or events united by some form of regular interaction or interdependence in the accomplishment of a specified outcome.

Why should we be concerned with all the fuss and bother required to develop an instructional system? Is not our teaching at the college level good? Frankly, no! at least not good enough. For example, there are other ways to teach than just the lecture method. Yet, it can be readily documented that by far the most frequently employed instructional methodology at the college level is that of lecturing. Why is this? At least two reasons readily come to mind: (1) lecturing is probably the most natural methodology for professors to employ; particularly since most professors have never received any formal teaching training; and (2) contingent upon the first reason, continuing increases in college undergraduate enrollments force the pedegogically untrained to persist in the lecture. Yet, the very fact of increasing numbers of college students results in reduced time available to the individual student and argues strongly for the need of more precisely organized instruction. The only way to break out of the lecture lock step and improve instruction is to systematically develop instructional systems that are valid for the purposes intended. This requires that all methods, materials and procedures be clearly identified, systematically developed, and refined until they effectively and efficiently achieve what they were intended to achieve.

Our present knowledge and skill in the "science" of developing instructional systems is admittedly limited, however, we have moved past the initial trial-and-error stages and are able to at

least prescribe a process that works, as Dr. Gabriel Ofiesh, Director of the Center for Educational Technology at the Catholic University of America, recently said, we have developed learning technology to the point "where we can start shooting with a bullet instead of a shotgun." (1967, p. 26)

In this section of the manual discussion we'll center on the important stages and procedures required for developing and validating an instr   onal system. These stages have emerged from developmental research over the past five years at Teaching Research and other research centers. The general process is a systems approach and requir s specific identification and specification of all elements of the instructional system along with systematic try-out, analysis and modifications until the desired outcomes are achieved. The flow chart presented in Figure 1 shows the major stages in the system.

Notice that stage 1 identifying behavioral objectives, was discussed in Section II of this manual and stages 2, 4, 5, and 6 were presented in Section III. Stages 3, 8b, and 9b, dealing with measurement and test construction, is discussed in Section IV. Stages 7, 8, 9, 10, and 11 are specifically concerned with the process of building and validating an instructional system and are discussed in detail in this section. Stage 7 is concerned with the actual translation of specifications into the instructional product and involves content specialists, behavioral scientists and media experts; try-out of the instructional product to determine the effectiveness of the new system in the central issue of stage 8; In stage 9, analysis of the feedback from the try-out are discussed; stage 10 concerns the use of the analysis for making necessary modifications to the new instructional system; and recycling of this whole process to achieve continued refinement of the product occurs in stage 11.

The building and validation stages can be considered culminating activities to the total instructional system development; therefore in order for the reader to more fully understand the relationships among all stages in the Systems Design in Figure 1, a brief review will first be made of the other stages and their importance to the stages in this section.

Stage (1): The first stage in the development of an instructional system is to <u>specify</u> the <u>behavioral objectives.</u> First, test your memory on what you learned in Section II. In writing behavioral objectives, each statement should contain descriptions of the following elements: (Write in answers) 1)____________, 2)___________, 3)___________, and 4)___________. (Turn to page V-7 to check your answers.

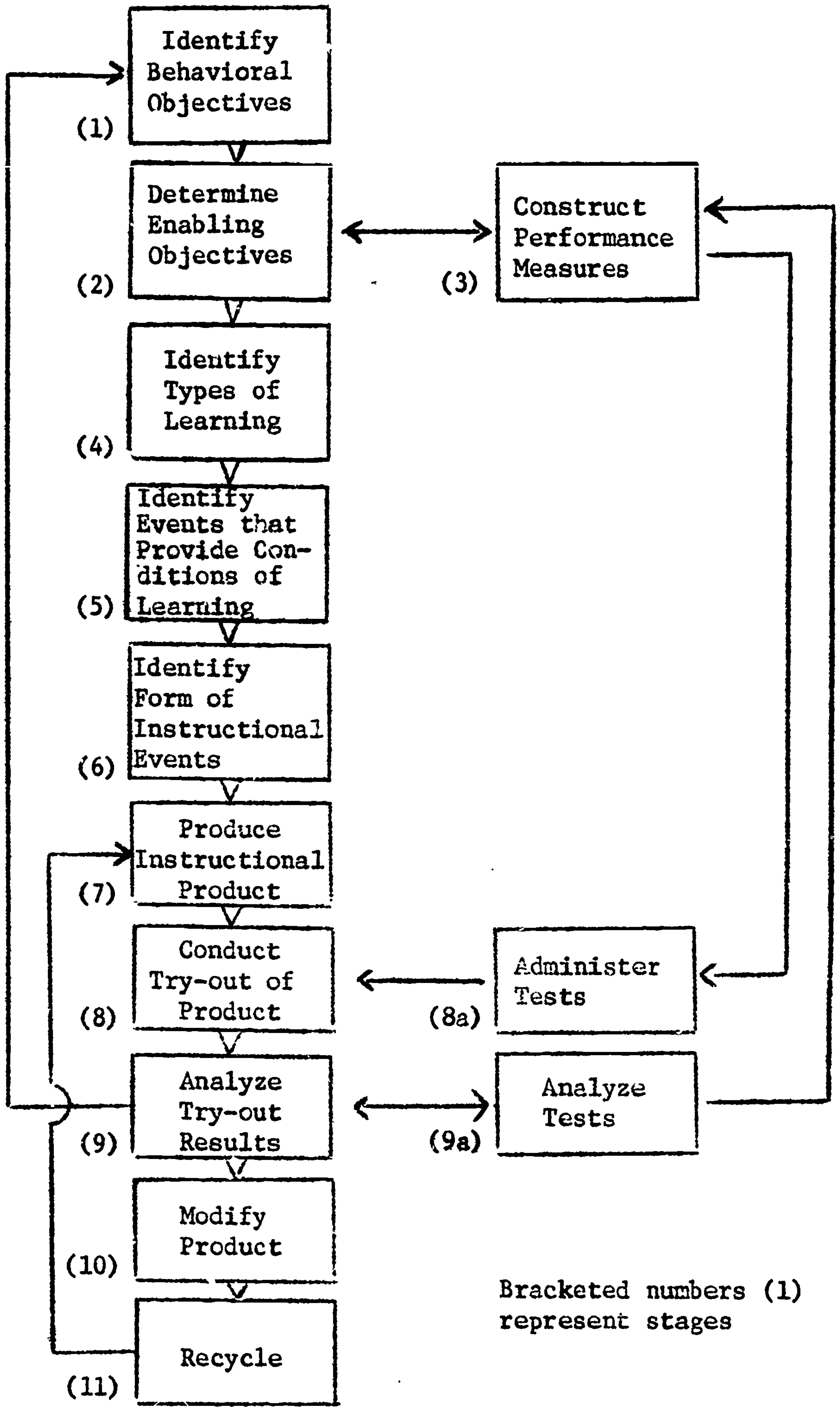


Figure 1. Major Stages in System Design for Developing Validated Instructional Systems

If we are to achieve precise learning outcomes from the instructional system that we develop, we must know very precisely what changes in behavior we expect of the learner and have a standard against which we can check to determine whether indeed he has actually gotten there. This is the place of behavioral objectives. Behavioral objectives are the very corner stone upon which the systematic development of an instructional system rests; and without them we need go no further.

Stage (2): The next stage in the development of an instructional system is to determine the enabling objectives. Test your recall again. In section III you learned that enabling objectives consist of ______________, ______________, ______________, etc. that enable the student to _______ the specified ______________. (Confirm your answers on Page V-7).

We must make certain assumptions regarding what the learner must know if he is to achieve the stated terminal objectives. These assumptions are based on empirical evidence, research, theory, and many times a seat-of-the-pants logic. They become the blue print specs that guide us in the development of the instructional system.

Stage (3): At this point in the development of an instructional system it is necessary to construct performance measures. Detailed instruction on this concept are presented in Section IV. However, a few important points will be made here to stress the important place that performance measurement holds in instructional system development.

Criterion measures are obviously necessary for determining whether the learner meets or exceeds the level of performance expected of him for each behavioral objective. If this was the only use for performance measures there would only be minimal need for this stage to be included in the system design and no need for it to be located opposite stage (2), determining enabling objectives. It could just as well be tied directly to stage (1), determining behavioral objectives. However, two very important additional functions are served by performance measures in instructional systems development that argue strongly for performance test construction to be tied to stage (2), determining enabling objectives. First, we need a diagnostic means for determining the validity of various steps of the instructional system. By developing performance tests from enabling objectives we produce a kind of grid like structure for assessing all parts and points of the instructional system to determine where weaknesses exist in the system.

This notion expands the concept of just listing performance tests to assess learner behavior, and permits inclusion of performance measurement as a means of assessing all elements of the instructional system.

Second, at our present level of sophistication, one of the weakest links in instructional systems development concerns the assumptions underlying the determination of enabling objectives. For example, as subject-matter properties to be generalized and discriminated become more subtle, such as the musical concept of early or late Beethoven, the instructional process become more complicated. A major problem in teaching such subtle and complex concepts is in the analysis and definition of subject-matter properties. Such analysis becomes more difficult when sematic confusion exists and where there is disagreement among experts. Assumptions regarding what elements of knowledge or what schedules of reinforcements, are required to enable the learner to attain the desired terminal behaviors, for example, are apt to be faulty and must be tested. Performance measures designed to test assumptions upon which enabling objectives have been determined improve the chances of making the steps in the instructional system valid.

Stage (4): The fourth stage in our systems design flow chart is to <u>identify</u> the <u>types</u> of <u>learning</u> represented for each objective. In Section III you learned that Gagne distinguishes eight types of learning. Three of these types of learning were considered more important than the other types for college level instruction. Determine for the three examples that follow which type of learning is represented. (Write your answers in the appropriate space; then check your answers on page V-7).

a. The high school third baseman fields the ball from the hitter, looks at the runner on second base, then throws the runner out at first base. ____________________

b. The college student replans his course schedule as a result of dropping one class and taking a new one. ____________________

c. The teacher plays a recording of a Mozart symphony. The student raises his hand when the music changes from a major to a minor mode. ____________________

Once the types of learning represented in each objective have been identified then strategies by which they can be made manifest can be handled systematically in developing the new instructional system. This narrows down the possible array of alternatives from which to choose in developing the product.

Stage (5): When the types of learning have been determined, then stage five becomes necessary: i.e., to identify the events that provide the conditions of learning to occur for each objective. Can you recall the first four steps for instruction of principle learning presented in Chapter III? Summarize each step below and check your answer on page V-7.

Step 1.____

____.

Step 2.____

____.

Step 3.____

____.

Step 4.____

____.

By identifying the events that permit the conditions of learning to take place, additional closure is reached regarding the type of strategy to be used in developing various parts of the instructional system.

Answers to questions:

Stage (1)

In writing behavioral objectives, each statement should contain descriptions of the (1) Audience, (2) Behavior, (3) Conditions, and (4) Degree.

Stage (2)

Enabling objectives consist of component actions, knowledges, skills, etc. that enable the student to attain the specified behavioral (or terminal) objectives.

Stage (4)

The types of learning represented in the examples are:

a. Principle learning
b. Problem solving
c. Concept learning

Stage (5)

Four steps required for instruction of principle learning:

Step 1. Inform the learner about the form of the performance to be expected when learning is complete.

Step 2. Question the learner in a way that requires recall of the previously learned concepts making up the principle.

Step 3. Use cues to lead the learner to put the principle together in the proper order.

Step 4. Using a question, ask the learner to "demonstrate" one or more concrete instances of the principle.

Stage (6): The sixth stage of the system design is to <u>identify</u> the <u>form</u> of the <u>instructional event</u>. That is, specify what form (verbal, visual, etc.) is to be used for the various types of learning identified to transmit the content to the learner. Perhaps this is our weakest link in translating specifications into products. We do have some research evidence to guide the use of audio and visual elements but this evidence is still very thin. We must rely mostly on seat-of-the-pants logic and empirical evidence of pilot tryouts to determine the "best" form to use; and even then time and cost prohibit any large scale trial.

All the stages in the system design thus far discussed are essential to actually building the system. Methods for gathering input to develop the instructional system and for determining required output of the system have been described in detail. Input is to be identified in terms of prerequisites, enabling objectives and conditions for learning; required output takes the form of specified behavioral objectives. In addition, the means of identifying instructional strategies along with supporting media forms are to be described. It can be assumed that the content is selected and will be sequences and that special equipment can be determined. Criterion measures, both intermediate and end of course, are to be prepared. Now we must produce the parts of the instructional system systematically assemble them, try them out with learners, appraise their effects, analyze the outcomes, make necessary modifications and recycle the whole process until we achieve the desired outcomes.

The development of a complex instructional system necessarily includes interactions among the learners, the teacher, and all other elements which constitute the system. However, if we are to be successful in developing such a complex instructional system we must give close attention to the validity of the system. In the past the failure to test the validity of instructional planning has limited the effectiveness of instruction. Only as all aspects of an instructional system are systematically tested can feedback begin to direct our attention to portions that are weak. But equally important to testing is an organized instructional system that can be replicated, thereby permitting the validation process to occur.

The key elements of a validated learning system are the following:

1). The end point is learning not teaching, all too often teachers place so much emphasis on the teaching role that they tend to lose sight of whether or not learning took place.

2). Outcomes are achieved through a system. The system includes all of the materials, equipment, facilities, procedures, texts, program schedules, maintenance, and personnel required to produce the end results.

3). The entire process must be validated. Only when the system is validated can we have any confidence that we are achieving what we set out to accomplish. Understand that the emphasis here is in testing the pedagogy, not the student.

Actually, the development of an instructional system is more akin to an "engineering process" in learning. The instructional specifications or "blueprints" must be translated into actual materials objects and viable procedures by an instructional engineer or "contractor;" who in turn must test out, analyze and modify the newly developed system until the finished product actually produces the type of performance called for in the original statement of objectives (learners exhibit specified terminal behaviors **they have learned**).

Stage (7): The seventh stage in developing the instructional system is to produce the instructional product. This becomes the first stage of actually creating the instructional system--systematically following the instructional blueprints or specifications, to give the system substance, form, and order. The process is one of translating written statements (specifications) into prototype of materials, equipment designates, and instructional sequences. The process is analogous to the Boeing structural engineer creating the complex 727 system from a set of blueprints and specifications. In both cases the degree of system integrity determines the extent of functional success--obviously more noticeable if lacking in the 727 system but no less important in the instructional system.

In creating the instructional product it is essential to systematically follow the instructional specifications. Content must be written into statements or translated into visual and/or audio forms, and arranged into sequences that will presumably accomplish desired behavioral changes. Media forms must be determined for each aspect of the content; i.e. whether to use printed forms, slides, overhead transparencies, audio tapes, 8mm loops, 16mm movies, various combinations of the above, etc. Formats for each selected media form must be established; i.e., whether printed statements are to be hand lettered or typed and enlarged; whether to photograph real objects, caricatures, or abstract symbols; whether to use black and white or color. Strategies for sequencing the content and media and for involving the learner and the teacher must be decided upon which will presumably be most effective in producing behavior changes and also interesting to the learner. Decisions must be made detailing exactly what content elements go first and which should come next; transitions between elements must be specified; what actions and routines the learners should experience must be planned; how the teacher is to interact in the learning situation must be determined.

In addition to the above, the preparation of a teacher's manual or guide is important. If the instructional system is to produce outcomes as designed, the teacher using it must be provided a detailed set of instructions explaining the exact purpose of the system and how it should be employed. Simple systems, such as a self-instructional program, usually have minimal teacher involvement during instruction, however, the manner of introduction and follow-up have considerable importance to its over-all effectiveness. The more complex the system the more crucial the role of the teacher.

During the product development it is essential to maintain close liaison between content specialists, media specialists and

behavioral scientists. To the extent that each knows what the other is doing greater developmental continuity will result. For example, if visuals are specified to represent certain content portions, considerable headway can be made if meetings are held with a graphics artist, content specialist and learning specialist during which suggested sketches can be quickly prepared, and jointly viewed and interpreted. Similarly, if a media specialist can advise on media forms, graphics work can be prepared permitting more efficient transformation into the final product.

Even with specification given, numerous decisions must be made during the development. For example, although content has been defined in the specifications, exactly what combination of words should be used, what language level represented, what format for displays, which examples to use, what exercises to include, etc. Questions such as these must be answered in the actual development.

It is also not unlikely that many gaps will exist in the specifications which must be confronted by the developer. Regardless of the detail in specification preparation, it must be realized that still lacking is the sophisticated knowledge necessary to prescribe, in the enabling objectives, all the conditions, strategies, and forms required in the instructional system to produce optimum outcomes.

Several guidelines have emerged that assist the product developer in making decisions during development. The model presented in Figure 2 represents the total continuum of experience, from abstract to reality. This model is helpful in deciding what media to use and when.

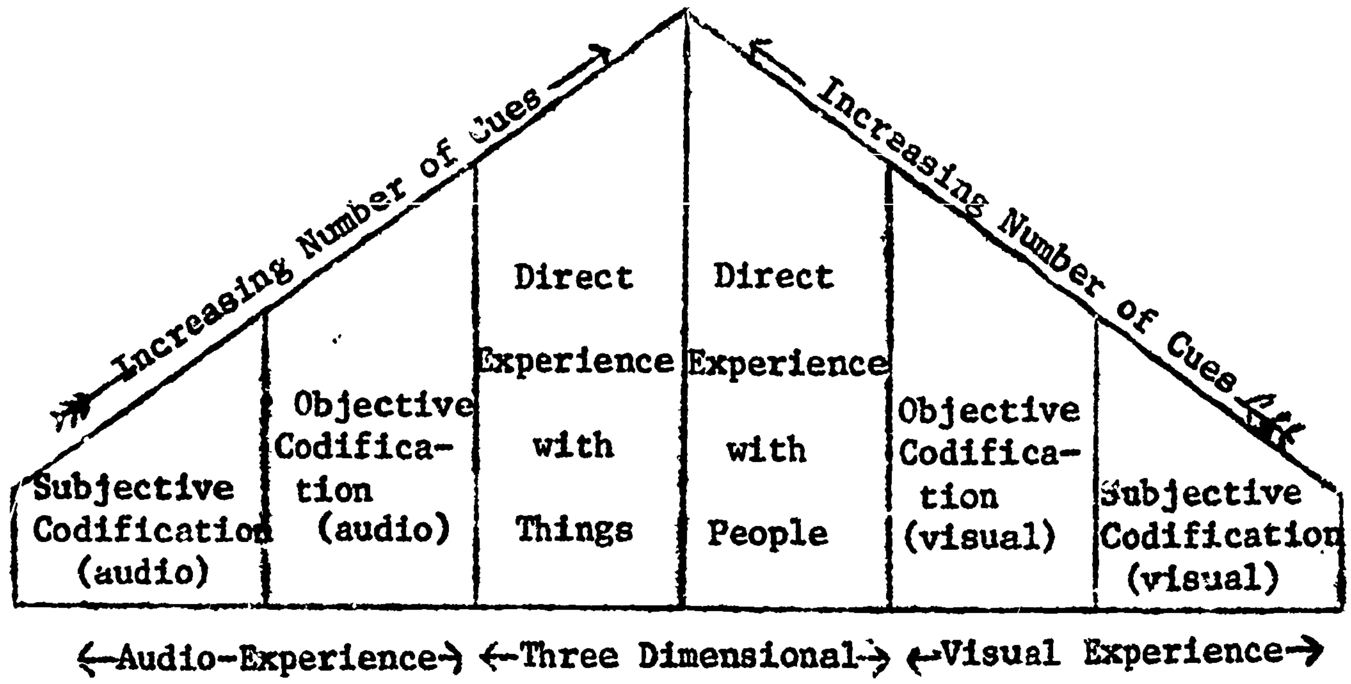


Figure 2. The Continuum of Experience

For clarification of Figure 2, each of the sections in the continuum model starting from the center and working alternatively toward the ends will be explained. First, direct experience with things. This section concerns real objects and ranges from the object itself to models representing the object. Objects are presented in three dimensional form and increase in number of representational cues from fewer in mock-ups to most in the object itself. The continuum of direct experiences with things is shown in figure 3.

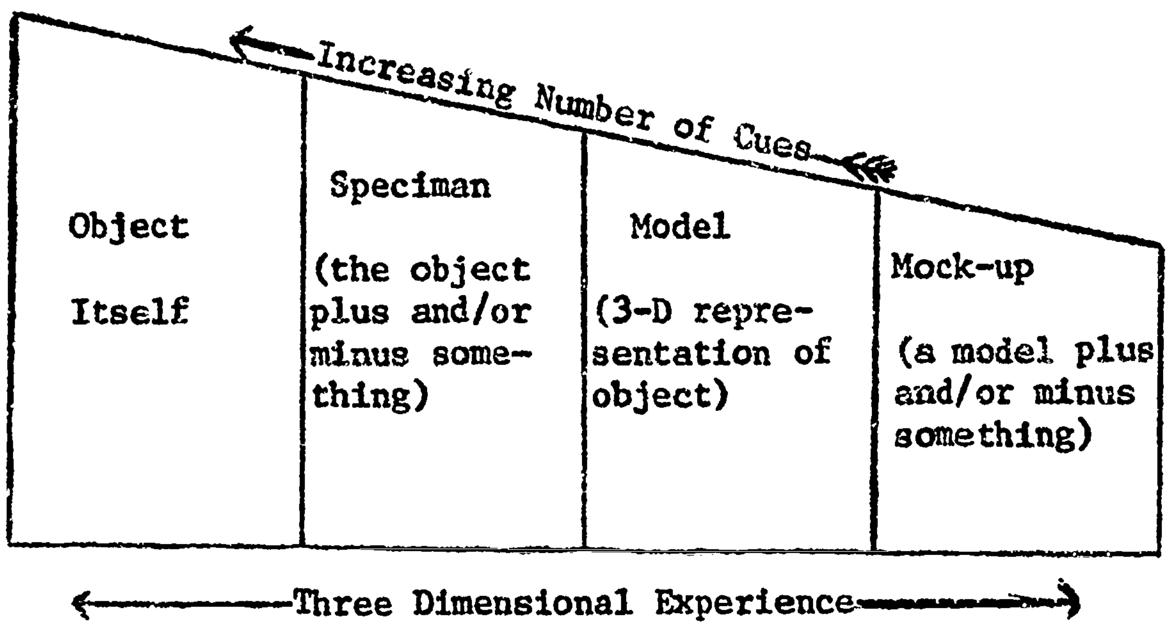


Figure 3. Direct Experience with Things Continuum

Direct experience with people consists of continuum ranging from authentic or real situations through pantomimes of situations to demonstrations of situations. These experiences are representational cues from fewer in demonstrations to most in authentic situations. Figure 4 contains the continuum of direct experiences with people.

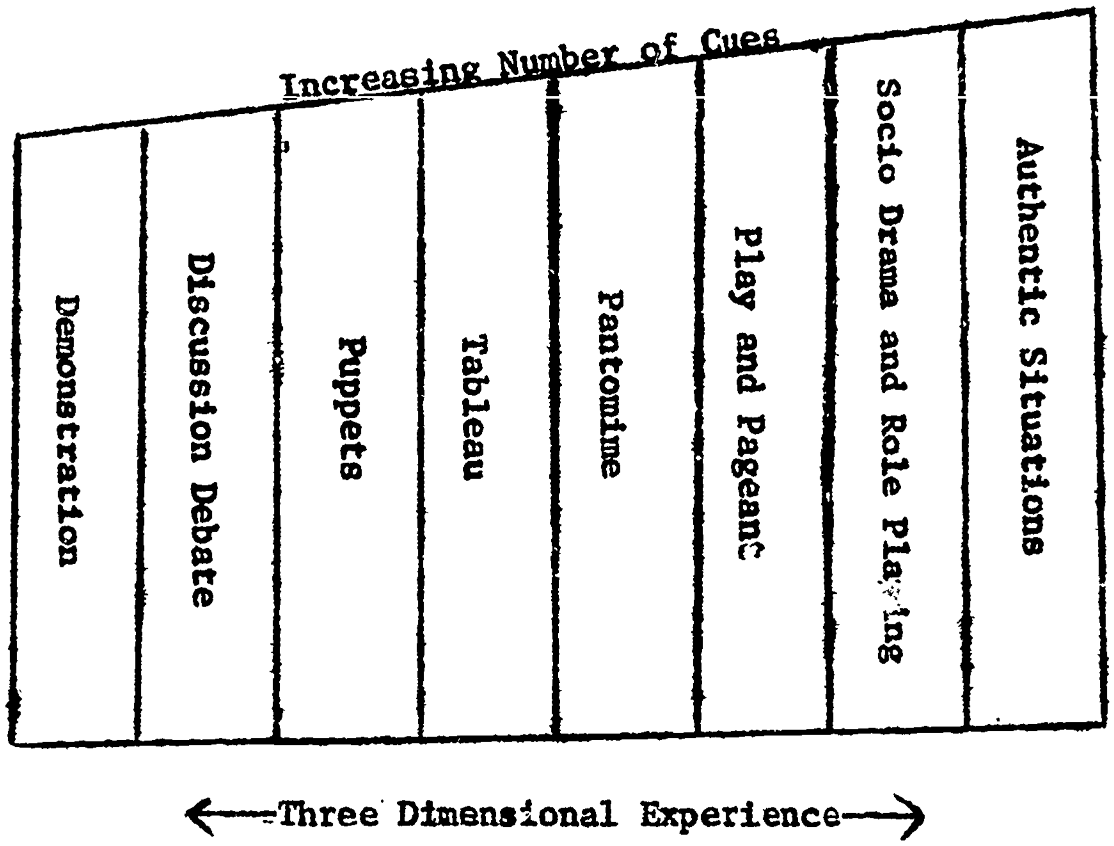


Figure 4. Direct Experience with People Continuum

The next two sections concern objective codification which is defined as: of or having to do with a known or perceived object as distinguished from something existing only in the mind of the subject or person thinking, hence, being, or regarded as being, independent of the mind; real; actual. Determined by and emphasizing the features and characteristics of the object, or thing dealt with rather than the thoughts, feelings, etc. of the artist, writer, or speaker; as an objective description, painting, etc. hence, without basis of prejudice; detached; impersonal. Objective codification with visual experiences are represented in two dimensional visual form and range from representational cartoons without color having fewer cues to motion pictures, with illusion

of 3 D, in color having most cues. Ordered elements representing <u>objective</u> <u>codification</u> with <u>visual</u> experiences are shown in Figure 5.

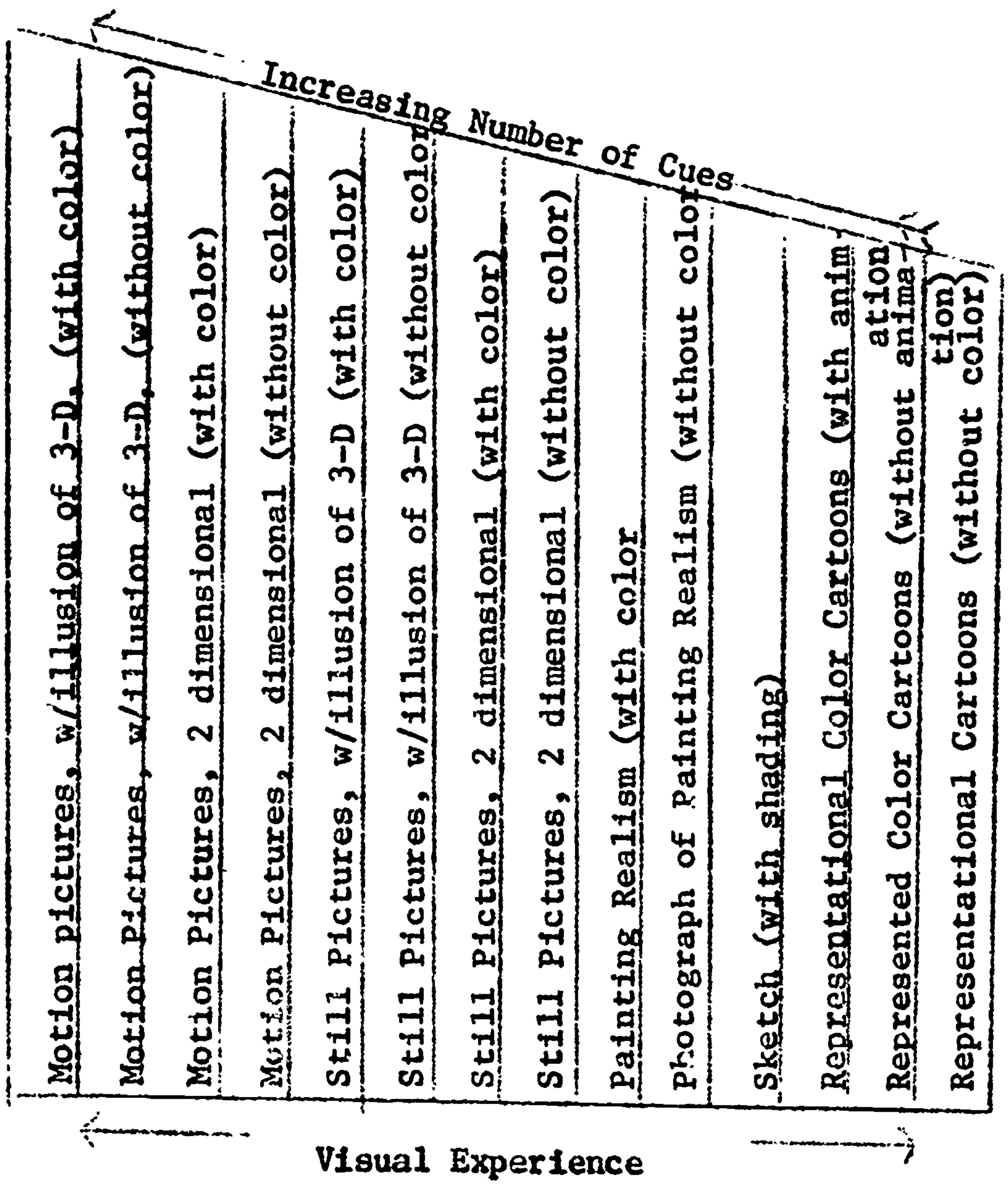


Figure 5. Objective Codification with visual Experience Continuum

The continuum representing objective codification employing audio experiences is presented in figure 6.

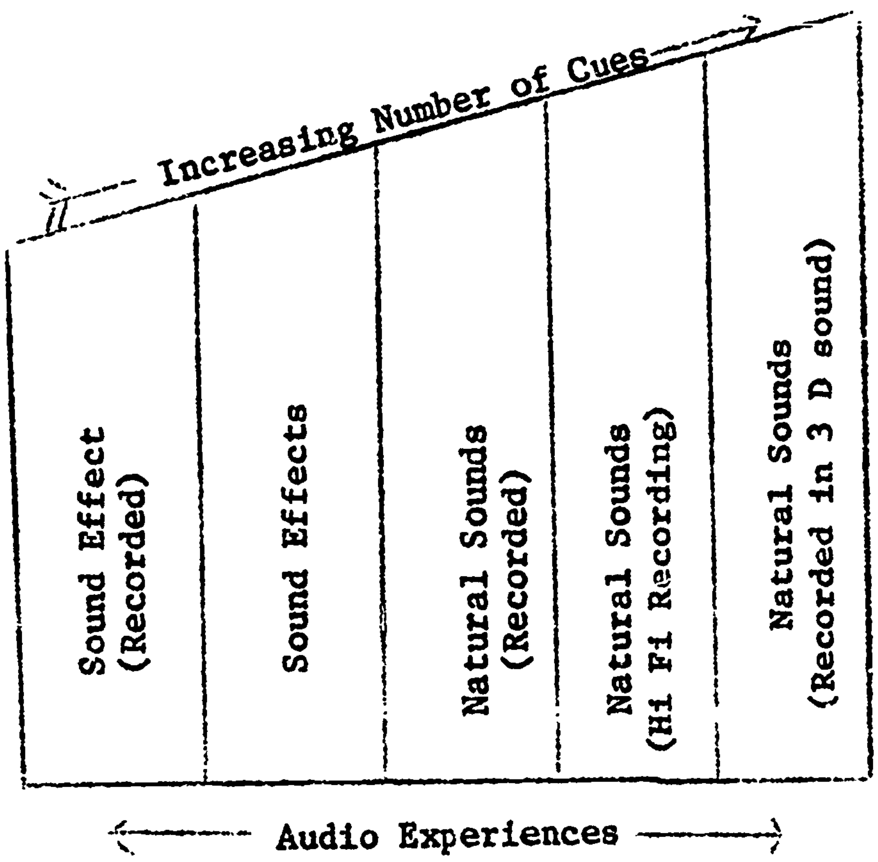


Figure 6. Objective Codification with Audio Experience Continuum

The tail end sections of figure 2 concern subjective codification which is defined as: of, affected by, or produced by the mind, or a particular state of mind; or resulting from the feelings or temperament of the subject, or person thinking, rather than the attributes of the object thought of: as, a subjective judgement. In Psychology: existing or originating within the observer's mind and hence incapable of being checked externally or varified by other persons. Subjective codification with visual experiences range from abbreviated symbols (shorthand, etc.) having fewer cues to symbolic cartoons having most cues. Figure 7 contains the con-

tinuum of subjective codification with visual experiences.

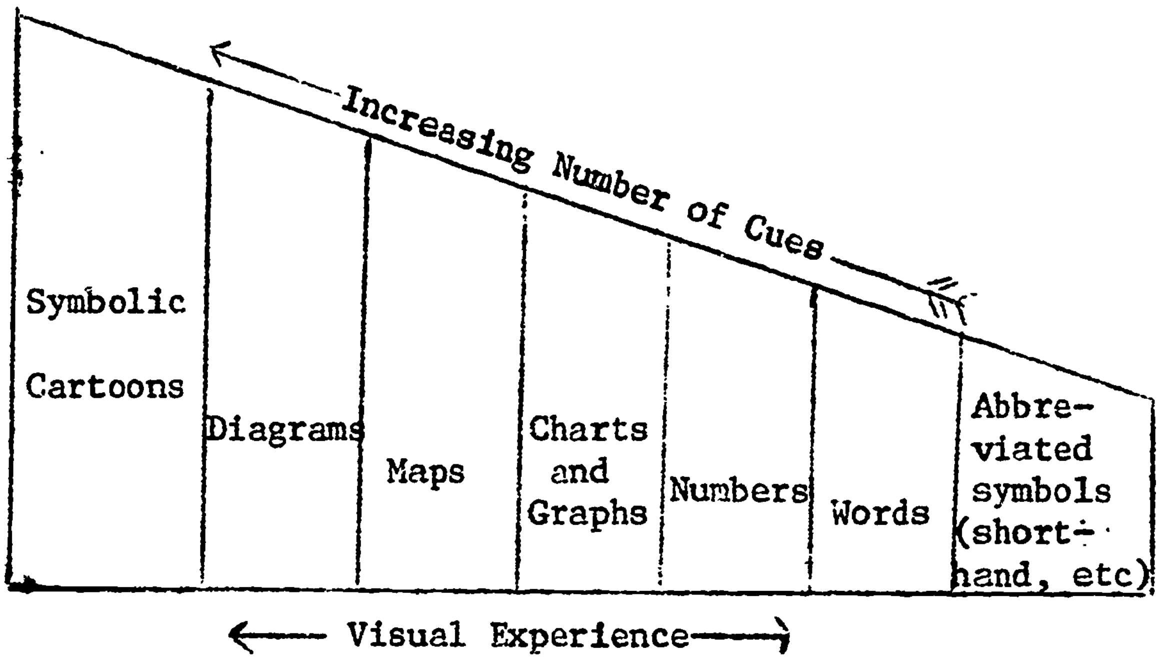


Figure 7. Subjective Codification with Visual Experience Continuum

The continuum of subjective codification with audio experiences is presented in Figure 8.

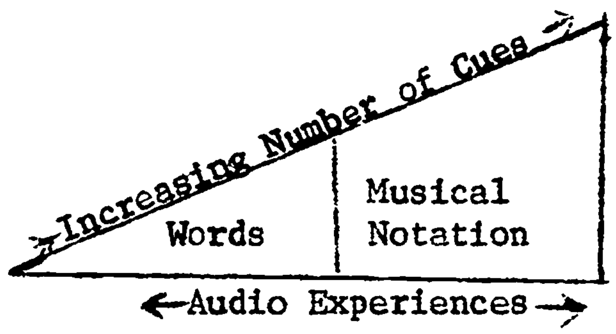


Figure 8. Subjective Codification with Audio Experience Continuum

The following general guidelines derive from the continuum of experience model and are helpful in making decisions about the form of the instructional system.

1. If the object or event to be taught is available and cues can be perceived, then use direct experience. For example, if

youngsters are being taught about the post offices and one is available in the immediate area, it is preferable to take the youngsters to the building rather than employing symbolic representations.

2. If the object or event to be taught is not available and cues are not readily perceived, use pictoral representation. For example, underwater phenomena are generally unavailable for direct classroom instruction, therefore objectively codify them (present them by visual means).

3. If the object or event to be taught makes vibrations and these are significant cues and are available, use natural sounds. For example, the sharp rattling sounds emitted from the tail of a rattlesnake are more effectively taught by demonstrating the natural sound.

4. If the object or event to be taught makes vibrations and these are significant cues and are not available, use auditory representations. For example, heartbeat sounds symptomatic of various maladies are readily available through objective codification (tape recordings).

5. If the object or event to be taught is psychological in nature, that is, no iconic (image) representation is possible, use audio and/or visual symbols. For example, since no one has ever seen an atom, subjective codification must be employed in instruction.

6. If the object or event to be taught is confusing, complex, and/or hidden use symbolic representation. For example, subjective codification (visual and audio experiences) must be used to transmit musical ideas.

Stage (8): The next stage in instructional systems development is to conduct try-outs of the product. This is basically a quality control measure built into the developmental process. The concern of this stage is to subject the newly developed system to try-outs with appropriate target populations to determine whether the content, strategy and supporting facilities, equipment and materials of instruction actually do produce the desired changes in behavior. The primary purpose here is to obtain data from observations and other evaluations of the system such that weaknesses in the system can be identified and changed.

Although this stage implies end-of-development try-out, actually try-out must be ongoing throughout the entire development. If the instructional development process is not interrupted periodically to take account of what has already been completed, faults in the specifications or the production already developed are apt to be overlooked thereby resulting in considerable waste of time and money at a much later time. For example, as the content of the instructional system is developed into specific written statements, unless statements are checked for accuracy, tried out for language level, judged for interest level, etc. only limited confidence can be held that these elements will indeed be adequate at completion.

The term try-out tends to imply involvement only with appropriate type learners to determine if their behavior will be adequately changed. This interpretation is much too narrow and would be too slow and limited a process to produce all of the desired feedback. The term target population used above includes all possible sources to provide feedback of the system's effectiveness. For example, in addition to feedback from learners, written portions of the system should be submitted to an "unbiased" discipline expert to check accuracy of content. Unbiased is used here to mean some one who has not previously been involved in the new system development. Feedback from a learning specialist should be sought to check integrity of examples to learning types and the efficacy of the events employed to produce intended types of learning. Media specialists should check fidelity and quality of audio and visual elements; a "typical" teacher should react to the instructional design to identify awkward and/or impractical routines. The above are representative check areas and others should be engaged as the system suggests. To the extent that feedback can be obtained from various specialists prior to involvement with actual learners, the chances of obtaining maximum beneficial feedback in the try-out with learners is increased.

Extremely important to the try-out is selecting the individuals to cooperate. Care should be taken that they represent an appropriate level of competence or ability. Learners should be representative of the audience for whom the system is intended. However, attention should be given to locate individuals willing to take the time and "go to the bother" of conscientiously giving feedback of their trial efforts and who are more apt to give "honest" responses. Subject matter, learning, media and other specialists should be skilled and obviously competent to advise in their respective areas.

The learner must be thoroughly briefed on what behavior is expected of him during try-out. The emphasis should be placed on his seriously trying to learn from the materials. However, he should be prompted to verbalize about the places he finds confusing, too difficult, already known, etc. Further, he should be encouraged to comment if he finds dull and uninteresting areas.

Close observation must be maintained during the try out with learners to produce the required feedback. Whereas curriculum studies and other materials developments such as commercially prepared products, tend to take large segments of the completed materials into the classroom for testing, the process here advocated is to try-out and test small segments of the unfinished system with a single learner at a time. The learner is instructed to identify any elements of the materials that are confusing or difficult to understand, such as words, phrases, pictures, symbols, test items, etc. The try-out monitor makes careful record of the location of all such identifications along with any other significant occurrances during the try-out; i.e., puzzled expressions, evidence of being bored, undue time required to complete various sections, etc. If the monitor suspects the learner is having difficulties that he is not expressing he should interrupt the instruction and tactfully question the learner. Care must be taken in questioning not to create anxiety in the learner. If the subject becomes anxious the value of that try-out is greatly reduced and it would be better to discount the effort and reschedule another try-out with a different learner. It is important to conduct try-outs with as much instructional realism as possible, i.e., present the learner with intact instruction beyond the designed teaching routines.

As large segments of the new instructional system are developed, more emphasis is given to learner try-outs and less to specialist feedback. The emphasis shifts from individual to small group learner try-out. Less close observation to the individual is maintained and more reliance is placed on test results.

To the extent that tests have been developed for sub-portions of the new system and their validity and reliability are judged sufficiently high, such tests become diagnostic tools for determining weaknesses in the materials.

Certain details accompanying try-out activities are important not to overlook. Clearance for using try-out subjects with appropriate administrative level authorities prior to try-outs will prevent possible subsequent misunderstandings and will greatly improve the chances of obtaining additional future try-out subjects. In some instances providing transportation for subjects will avoid time lost and will preserve schedules. Monitors must be trained prior to try-outs in what to observe and how to behave during try-outs. In situations involving total classroom try-outs, detailed instructions to cooperating teachers specifying all required aspects of the try-out are essential.

When all development of the instructional system is completed, field testing in context is the necessary final stage of try-out. By this time, segmented and small scale try-outs should be completed and necessary revisions or modifications to the system processed. At this point it is essential to try the total instructional system in the context of an actual classroom with all its normal contingencies (expected or otherwise), to see if the expected changes in learner behaviors occur. The system should be sufficiently complete in all parts--instructional materials, teacher guide, tests--that when handed intact to a teacher, classroom instruction can be readily initated. If, in context, the system brings the expected proportion of learners to criterion within the allotted amount of time then development has been successfully completed.

Now consider a brief example in which the instructional system is at the earlier stages of development. Presume that development is underway on instructional materials designed to teach 8th graders to use the Readers' Guide. Specifications call for visuals in the form of overhead transparencies designed to present concepts of the Readers' Guide referencing system and to be accompanied by student worksheets containing exploratory statements and elements requiring application. Several parts have been developed: visuals that replicate specific Readers' Guide reference entries and work sheet items keyed to the visuals that define various symbols and abbreviations used in the entries and explain their relationships. One of the prepared visuals contains

the following entry:

Aeronautics: Mid-air payoff. il Time 93:62+
Jl 21 '67

The accompanying Student Worksheet had the following exercise in conjunction with the visual:

> On the screen is a sample of an entry from the Readers' Guide. In the spaces that follow, copy the following information: Title of article, author, type of illustration, name of magazine, volume, page, month, day, and year.

A Teacher's Guide was also a part of the system and had explanatory notes for the teacher. The following note was prepared for the visual.

> Using the overlay, discuss (1) author and title, (2) illustrations, (3), name of magazine, (4) volume number and pages, and (5) date.

Now put yourself in a hypothetical try-out situation with a single learner. Regard the above visual element, worksheet statement, and teacher note as the segment you are to try-out. Although a typical try-out segment would involve more than a single item series, such as above, for the sake of working through a problem example, assume that the above elements are a larger segment of an instructional system. In the spaces below identify briefly (1) what details you would consider in arranging the try-out and (2) how you would conduct the try-out; also (3) guess at what effects you might observe during the try-out regarding (a) what comments the learner might make (b) what reactions you might observe in the learner, and (c) what reactions you might observe from the Teachers' Guide.

(1). Arranging the try out:

(2). Conducting try out:

(3). a. Possible learner comments:

b. Possible learner reactions:

c. Possible reaction to the Teachers' Guide

You might wish to compare your answers with those suggested on page IV. Differences can be discussed during the institute group sessions.

Suggested answers to hypothetical try-out situation:

ANSWER

1) Arranging the try-out. It will be necessary to prepare all necessary try-out facilities including A-V equipment; a "typical" 8th grade student must be obtained from the local schools to act as try-out subject ; arrangements for using the student are made through the superintendent; arrangements for car transportation and a schedule of pick-up and return of student are made; extra copies of all instructional materials and tests are produced.

ANSWER

2) Conducting the try-out. The learner is briefed on how he is to behave during the try-out; instruction is provided, carefully following all prescribed pedegogical details; notes of all student comments and responses are written into the materials at the precise point of occurrence; at conclusion of instruction, criterion measures are administered.

ANSWER

3) a. Possible learner comments. The learner is apt to state that he cannot find the author's name from the visual; it is possible he will state that he does not understand what to write in the worksheet.

b. Possible learner reactions. The learner might be observed copying into his worksheet all details from the visual without any apparent understanding; or the learner might appear confused and not be writing anything.

Cont'd 3

c. <u>Possible reactions to the Teachers' Guide</u>. The guide materials do not parallel the student materials. No explanation is given for extra details such as author/no author, volume number, + sign.

Stage (9): Another quality control point in developing an instructional system is to <u>analyze</u> the <u>try-out</u> of the <u>product</u>. The purpose of this stage is to determine whether the instructional system attains its objectives. Two primary uses are made of this analysis: 1) to determine what portions of the new system are faulty or not adequate and require modification, and (2) to determine which specifications were unrealistically and/or improperly established and should be changed.

Analysis of the try out of a new instructional system is not nearly so precise as that resulting from the customary experimental research: The principal difference centers around controls. Where experimental studies place considerable emphasis on appropriate controls, developmental research of new instructional systems, on the other hand, seldom offers any basis for controls. The new development is just that, new! No suitable comparable thing exists against which it can be compared. Similarly, no substantial basis exists against which to control.

Analysis must be maintained on an objective rather than comparative basis. That is, how far does the system go toward achieving the ends for which it was developed rather than how much better did the system perform compared to some other system. When a new system has been sufficiently developed and refined and in use over a period of time such that its accomplishments can be predicted, then comparative analyses with another established system are appropriate. When the system is still new and essentially untried the only basis for judgment lies in how well it attains the objectives for which it was created.

Other developmental research efforts also employ the objective oriented analysis for making judgments. For example, much of the performance of the Boeing 727 in wind tunnel try-outs was analyzed against new objective standards. Although certain factors were common to other older systems, the complex combination of old and new had to be very rigorously tested against newly established standards. The only basis for acceptance was successful attainment of the standards. If a portion of the 727 system failed in the test, careful analysis was made of what contributed to the failure so that modifications could

be made. The analysis often could not fix the exact cause of the fault, however, it did pin-point the area of weakness and thereby contributed significantly to planning for modifications and subsequent new trials.

How, then, does one proceed in objective analysis? It is tempting to say all is fair in love and new instructional system analysis and anything goes. However something better must be said if the concepts of research and the "scientific method" are to be amplified. The difficulty is that the scientific method, in the form that it is generally regarded, other than the ideal of rigorousness, has little guidance to offer this form of developmental research. We still know too little about how best to analyze the results of systems try-outs to prescribe any precise methodology. Literally, any and all means for analyzing results that meet the concept of rigorousness are currently being employed.

Generally speaking, analysis must closely parallel try-out activities. That is, as small segments of the instructional system are tried out with various specialists and individual learners, analysis of these try-outs must occur in order that necessary modifications to the segments and next steps in the system development can be made.

Feedback from specialists can occur in various forms depending upon the arrangements made. Discipline specialists will usually prefer to study the materials content privately making editorial notes of both a technical and stylistic nature. The developer usually reviews these notes with the specialist for clarification before analyzing their implications. Media specialists will frequently prefer to judge the materials in concert with the developer to be sure of interpretations. Criticisms in these instances are noted in detail by the developer and through this interaction take the form of analysis. Try-out with a typical teacher is usually jointly managed with the developer present. As in the situation with the media specialist, the analysis results from the interaction of the two.

Analysis of small segment try-outs with learners depends upon the nature of the observations employed and the form which they are recorded. In many instances, the very fact that learners express having difficulty with a particular portion of the new system will cause the developer to view that portion in different perspective. What appear to be obvious verbal flaws or sequences sometimes almost "spring out" at the analyst as a result of such learner feedback. For example, analyze the Readers' Guide

elements you hypothetically tried out in stage 8, Page V-21. It probably became immediately clear to the reader as a result of the learner's statement that he could not find the author's name in the visual that the visual did not include an author, yet the author was called for in the worksheet question. This form of analysis points to obvicus flaws in the development that were somehow overlooked in handling the plethora of details and lead directly to revisions.

Other forms of analysis are often less obvious. For example, perhaps the only evidence of weakness in an instructional segment is that each try-out subject has not learned a particular bit of information coming from a specific part of the instruction. However, no obvious flaws, as above, are evident in the materials. Careful study of the segment might reveal, for example, that double meanings are possible in the choice of words included and have led learners to an erroneous interpretation, or that particular sequential elements have combined to produce conditions of retroactive inhibition.

Look again at the Readers' Guide example on Page V-21 and make a further analysis. Give consideration in your analysis to both the visual and worksheet portions and suggested answer 3 given on page V-22. Briefly summarize your analysis in the space that follows, then check your answer with that given on page V-26.

<u>Analysis:</u>

ANSWER

Analysis: The visual gives too much information at one time; perhaps separation of elements would focus learner attention more precisely to the desired portions. Worksheet format does not permit the learner any routine order for writing down answers. This format combined with the too much detail in the visual results in learner confusion. Also, lack of author's name in the visual confuses the learner.

A more formal means of analysis than that described above occurs with criterion testing. Considerable attention to validity and reliability of the tests are necessary. Gagne, in describing research dealing with attainment of learning sets relevant to terminal objectives, actually demonstrates one example of how to analyze the try out of an instructional program using criterion measures (1961, p. 10). He first developed a hierarcy of learning sets, similar to that displayed in section III of this manual, by employing the question "what would an individual have to know how to do in order to achieve successful performance of this particular class of task, assuming he were given only instructions?" Five major levels of learning sets were thus formed, two levels, III and IV, being sub-divided once each.

In testing the measure of transfer among learning sets Gagne analyzed four possible empirical relationships for passing and failing relevant higher-lower learning set combinations. These four were as follows: (1) ++: indicating positive transfer from lower to adjacent higher learning sets; (2)--: indicating failure of lower level set followed by failure of adjacent higher level set; (3) +-: indicating failure of lower level set followed by passing of adjacent higher level set; and (4) -+: indicating passing of a lower level set followed by failure of adjacent higher level set.

The patterns of pass-fail relationships on a criterion test administered following the learning program are presented in Table 1. (Gagne, 1961, P. 1D)

Table 1

Pass-Fail Relationships of Achievement Between Adjacent Lower and Higher Level Relevant Learning Sets (N=118)

Transfer to Learning Set	Frequency of pass-fail Pattern--Higher, Lower ++ (1)	-- (2)	+- (3)	-+ (4)
IV2 from IVA1	110	0	0	8
IV5 from IVA3	113	0	0	5
IIIA1 from IV2, IV3	85	0	7	26
IIIA2 from IV4,IV5,IV6	94	5	10	9
III1 from IV1	45	9	1	63
III2 from IV3,IIIA1	68	30	6	14
III3 from IVA2,IV3	75	25	7	11
III4 from IIIA2	62	40	4	12
II1 from IV2,III2,III3	34	70	3	11
II2 from IVA2, III3	41	60	2	15
II3 from III4	37	72	3	6
II4 from III4	9	85	0	24
I1 from II1, II2	25	78	2	13
I2 from II2, II3	28	80	3	7
I3 from II3, II4	6	104	0	8

Although these results were primarily intended for other purposes they do demonstrate a good example of determining weak areas in a newly developed instructional system. Those interested in the interpretation of the total table are referred to Gagne's monograph.

In Table 1, the frequencies in the -+ column indicate the points in the instructional program at which greater or lesser effectiveness was attained. Large frequencies are interpreted as points at which the program was relatively ineffective, small frequencies the points at which learning was effective. For example, as shown in Table 1, progression from lower level learning set IV1 to adjacent higher level learning set III1 was failed

by 63 of the 118 learners. Examination showed that only one frame of the program was devoted to the content of set III1. Obviously insufficient instruction was provided these 63 learners in that one frame for them to fully understand the new information. Similar results pertained to learning sets II4 and IIA1 in which 24 and 26 failures respectively occurred after successes at the lower level set. Depending upon the level of errors acceptable by the developer, other frames in the program in this example would be subject to similar scrutiny. This form of analysis obviously relies very heavily upon a powerful criterion test--one that tests all aspects of the instruction system.

Equally as crucial to the analysis of the instructional material try-out is analysis of the measuring instruments. Only as test items are developed, tested and modified until valid and reliable for measuring changes in behavior resulting from the instructional system can there be confidence that desired learning outcomes are actually being achieved. This step is mentioned here only to give it proper continuity; further discussion of it will be given in section IV.

Stage (10): As try-out and analysis of the new instructional system progresses, the purpose of all this is to provide a systematic basis for <u>instructional material modification</u>. During this stage, results of the analysis are translated into changes in the product. In some instances the revisions will virtually be known to be an improvement, in others only additional try-out will so determine. For example, analysis such as the omission of the author's name in the Readers' Guide materials should prompt a modification in the visual quite certain to improve that aspect; however, revisions to the visual and worksheet based on the additional analysis cannot be made with such confidence. In both cases they must be tried out to make certain that the modifications are effective. This latter is a crucial point to the successful completion of an effective product.

Our Boeing 727 analogy might strengthen this point. For example, early test flight trials of the 727 revealed that at a certain angle of climb, one designed for takeoffs, the engines lost significant power; however up to that angle no power loss was sustained. Analysis determined that the air intake housed in the vertical tail structure was placed too low to receive sufficient air at the critical climb angle. Modifications were made placing the air intake higher in the tail structure. However, the 727 remained unacceptable until additional flight tests were completed at various angles of climb to determine whether

takeoff objectives were being met.

Modifications should be planned with the same team of specialists who contributed to the development. For example, consider the <u>Readers' Guide</u> elements introduced on page V-21. Analysis revealed that the visual contained too much information at one time. Methods of reducing the amount of information should be determined in conjunction with the media specialist.

What suggestions do you have for changing the visual? Summarize briefly in the space below the modifications you think would make the visual more effective. Check your answer with the one given on page V-30.

<u>Suggested Modification</u>:

ANSWER

Suggested modifications: The visual should be changed from a single transparency into a base transparency with four overlays. The base should contain the title: Aeronautics: Mid-air payoff; the first overlay to contain the abbreviation for illustration: il; the second overlay to contain the magazine name: Time; the third overlay to contain volume number and pages: 93:62+; and the fourth overlay to contain the date: Jl 21 '67. In this manner, information can be presented in smaller segments and more control is thereby available to direct the learner's attention to the specific element being presented.

In conjunction with the revisions recommended for the visual the analysis suggests that other changes be made to the Readers' Guide example of instructional materials. The worksheet example, from page V-21, should be modified into outline format with each bit of information separated as follows:

On the screen is a sample of an entry from the Readers' Guide. In the spaces that follow copy the information from the visual explaining the meaning of each of the following:

1. Title of article______________________________

2. Author (if given)_____________________________

3. Type of illustrations (if any)__________________

4. Name of magazine______________________________

5. Volume and Page number_________________________

6. Month, day (if given) and year__________________

It is further suggested that a second visual be prepared containing a Readers' Guide entry having an author and other symbal variations not contained in the first sample entry. In addition the Teacher's Guide shown on page V-21 should be modified to present the information to the teacher in a form that parallels the student worksheet form. Explanatory information should be included for the teacher regarding placement of author's initials, the fact that some entries are without author, use

of color separating volume and page, reason for volume numbers when magazines are bound, meaning of the +, and abbreviations used in the date.

Stage (11): The final stage of developing the new instructional system concerns the recycle of all product development stages. That is, returning to stage 7, producing the instructional product, and going through the cycle again. The purpose of this stage is to point out the importance of continuing with the development, testing, analysis and revision process until the desired outcome learner behaviors are achieved.

Obviously this total process is not one of a series of four discrete stages of which the second cannot be started until the first is completed, etc.; rather, it is a complex process of overlapping and combining. It is not unusual for different stages to be occurring together; i.e., a trade back and forth between production, try-out, etc., various segments of the system. It might be more efficient, for example, to develop several similar visual elements at the same time even though they are located in different parts of the system. Feedback regarding their fidelity and readability can be obtained before the instructional sequence to which they belong is ready for trial, thus pushing towards time and cost efficiency.

To the extent the developer can perserve in the development sycle until desired behaviors are achieved his product will improve; assuming, of course, that the desired learner behaviors are indeed desirable.

Bibliography

De Cecco, John P., (Ed) Educational Technology. New York: Holt, Rinehart, and Winston, 1964.

Glaser, Robert, "Implications of Training Research for Education," in Ernest R. Hilgard (Ed). Theories of Learning and Instruction. NSSE, 1964, Part 1.

Glaser, Robert, "Toward a Behavioral Science Base for Instructional Design," in Robert Glaser (Ed) Teaching Machines and Programmed Learning, II. NEA: Washington D.C., 1965.

Glaser, Robert (Ed) Training Research and Education. New York: John Wiley & ons, Inc., New York, 1962.

Gropper, George L. "Learning from Visuals: Some Behavioral Considerations," AV Communications Review, 14:37-69, Spring, 1966.

"Innovation in Education," Educational Technology. Washington D.C.: Aerospace Education Foundation, Spring, 1967.

Lange, Phil C. (Ed), Programed Instruction. NSSE, 1967, Part II.

Vandermeer, A.W., "Systems Analysis and Media--A Perspe tive," AV Communications Review, 12:292-301, Fall, 1964.

Tracey, William R., Flynn, Edward B. Jr., and Legere, C.L. John, "Systems Approach Gets Results," Training in Business and Industry, 4:17+, June, 1967.

VI - DESIGN

Evaluation Sheet

We need to know how well the ideas and issues in this manual are communicated to you. You are the test audience for this material. To remove or strengthen the weak spots, to retain or improve the strong ones (if any), an account of your learning experience as you read these sections is crucial. Use this sheet as you study and jot down your reactions. Don't be concerned with typographical errors. We're concerned with how the message is coming through.

SECTION #________ TITLE__________________________

Give topic, paragraph or sentence, page	Importance Rate 1 2 3 4 5 Low - High	Clarity Rate 1 2 3 4 5 Low - High	Suggested Improvements

Over-all Rating of the Section: Use a 5-point Scale

Important					Understandable				
1	2	3	4	5	1	2	3	4	5
Very Low				Very High	Very Low				Very High
(Circle one)					(Circle one)				

PLEASE PUT ANY ADDITIONAL COMMENTS ON BACK

SECTION VI
Research Design

Table of Contents

Section VI

Research Design

John Gordon, Jr.

Overview

This chapter deals with the fundamental logic and logistics of instructional research. When you have completed the session we hope you will be able to do a "first level" analysis of a research proposal and provide some constructive comments. We believe analysis of other proposals to be the first step toward constructing your own design, but unfortunately as all of us know one task is only prerequisite, not supplementary, to the other.

Experimental Design

The many definitions of experimental design in educational research range from the general plan of attack to the specific method of assigning the experimental treatment to the experimental unit. For the purposes of this institute we'll stay with the latter and be as specific as possible. The experimental treatment is defined as those conditions the experimenter has decided to manipulate in a specific way. This rather gross category is also referred to as the independent variable. The experimental unit in our case in instructional research is usually the student or that unit which is being acted upon by the conditions of the experimental treatment. There are also times when it becomes necessary to consider the entire class unit as an experimental unit. Using the entire class as an experimental unit is typical in instructional research in colleges. Another way of locating the experimental unit is to look for that which will be measured to determine whether or not the experimental treatment has been effective. These measures taken on the experimental unit are also called the dependent variables.

Again, experimental design is the "method of assigning" the experimental treatment to the experimental unit. The most common method is called random assignment, a typical example of which is drawing numbers or names from a hat. The fundamental assumption underlying this method is that every experimental unit has an equal opportunity of receiving the experimental treatment(s).

The usual procedure for random assignment is to use the table of random numbers that is found in most statistics and design texts. The procedures are spelled out in each text but generally: 1) allots numbers to the experimental units, 2) picks any number in the table, 3) continues down or across that row of numbers matching those that are in the table with those that have been allotted, and 4) put them in either the experimental group or other group. Another way of looking at the equal probability assumption of random assignment is to consider it a way of eliminating systematic bias. A <u>systematic bias</u> is defined as any condition, other than treatment conditions, either within the experimental unit or in the environment or context of the experimental unit that would be present for all members of one group and not another. This undesirable condition would also have to have a measurable affect upon the dependent variables or those measures of the experimental unit which you are hoping to change. Random assignment is considered, by most, as the optimal way of designing an experiment. Other methods are at best, poor substitutes. These methods, such as matching, color balancing, and statistical control, e.g. covariance analysis, all have fundamental problems. Such methods are considered more fully in design texts and will not be covered here.

Another basic distinction made in most discussions of design is that between univariate and multivariate experimentation. Univariate means that one particular condition is varied and one particular measurement thought to be associated with the condition is assessed. For example, a movie (experimental condition) may be shown to some students and not to others and a knowledge test (simple measure) administered after having witnessed the film. Multivariate experimentation is based on the notion of multiple causation which assumes that a combination of variables or experimental treatments acts upon a number of dependent variables or different measures taken of the experimental unit. For example, the movie may be shown to a certain kind of student and this combination of student and film influence the student in a number of ways, e.g., for both attitude change and knowledge content. Any combination or sequence of independent variables or treatments and dependent variables or measures is classified as multivariate experimentation. A piece of advice for the beginning instructional researcher would be to attempt a univariate form of experiment before considering the multivariate although the latter are generally considered to be much more valuable and representative of reality.

Your major task in this set of self-study materials will be

an analysis and critique of two proposals that are included in the latter part of the materials. The following is a check list to be used as a guide.

1. What is or are the experimental treatments?
2. What is or are the experimental units?
3. What method of assigning the treatments to the units has been used, in other words, what design?
4. Using the designs 1-6 from Campbell and Stanley, determine in which category the particular proposal design falls? (Determine the R's, X's and O's)

A more important guide, however, is the comprehensive chart prepared by Drs. Donald Campbell and Julian Stanley which appears in the Handbook of Research on Teaching. The chart and definitions are the crux of an excellent discussion which you may wish to read and refer to in the future. The chart is shown in Table 1.

Once you have judged what design category is being used you are then able to move directly to the chart to assess the sources of invalidity of the particular design. Starting with the strengths (+'s), you search out sentences in the proposal which show that the author has made note of these possible sources of invalidity and how this design overcomes them. If the author has not made note of the strengths of his design then he is in error and should be amended. More important, however, is the determination of the weakness(es) of the design. The author should emphasize these weaknesses and his attempts to control or compromise these sources of error. The argument should be clear and strong as to why certain areas or sources could not be controlled. For example, study of a short duration would not have the maturation effect as one of its concerns and the proposal writer should make note of this. In other words the argument should be clear, highlighting those major sources of invalidity controlled either by the design or other means peculiar to the experiment, and those minor sources not thought to be of importance or at least compromised.

It now becomes imperative that we look more closely at the sources of invalidity stressed by Campbell and Stanley in their classic article. A list of these sources and their definitions can be found in Table 2. But if, for some reason, these definitions are not understandable we will attempt a short synopsis at this point using common examples. Read their definition first and this example second, if needed.

Table 1**

Sources of Invalidity for Designs 1 Through 6

	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Morality	Interaction of Selection and Maturation etc.	Interaction of Testing & X	Interaction of Selection & X	Reactive Arrangements	Multiple-X Interference
Pre-Experimental Designs:												
1. One-Shot Case Study X 0	-	-				-	-			-		
2. One-Group Pretest-Post-test 0 X 0	-	-	-	-	?	+	+	-	-	-	?	
3. Static-Group Comparison X 0 0	+	?	+	+	+	-	-	-		-		
True Experimental Designs:												
4. Pretest-Post-test Control Group Design	+	+	+	+	+	+	+	+	-	?	?	
5. Solomon Four-Group Design R 0 X 0 R 0 0 R X 0 R 0	+	+	+	+	+	+	+	+	+	?	?	
6. Post-test Only Control Group Design R X 0 R 0	+	+	+	+	+	+	+	+	+	?	?	

Sources of Invalidity*

* Minus indicates a weakness; plus, a factor controlled; question, a source of concern; blank, an irrelevant factor. These are rough indicators only. Complex designs and qualifying conditions may alter the indicated effects.

** Abridged from Chapter 5 "Experimental and Quasi-experimental Designs for Research on Teaching." Handbook of Research on Teaching (1963).

Table 2

Sources of Invalidity

1. History, the specific events occurring between the first and second measurement in addition to the experimental treatment.

2. Maturation, processes within the respondents operating as a function of the passage of time per se (not specific to the particular events), including growing older, growing hungrier, growing more tired, and the like.

3. Testing, the effects of taking a test upon the scores of a second testing.

4. Instrumentation, in which changes in the calibration of a measuring instrument or changes in the observers or scorers used may produce changes in the obtained measurements.

5. Statistical regression, operating where groups have been selected on the basis of their extreme scores.

6. Biases resulting in differential selection of respondents for the comparison groups.

7. Experimental mortality, or differential loss of respondents from the comparison groups.

8. Selection-maturation interaction, etc., which in certain of the multiple-group quasi-experimental designs, such as Design 10, is confounded with, i.e., might be mistaken for, the effect of the experimental treatment.

9. The reactive or interaction effect of testing, in which a pretest might increase or decrease the respondent's sensitivity or responsiveness to the experimental treatment and thus make the results obtained for a pretested population unrepresentative of the effects of the experimental variable for the unpretested universe from which the experimental respondents were selected.

10. The interaction effects of selection biases and the experimental treatment.

11. Reactive effects of experimental arrangements, which would preclude generalization about the effect of the experimental treatment upon persons being exposed to it in nonexperimental settings.

12. Multiple-treatment interference, likely to occur whenever multiple treatments are applied to the same respondents, because the effects of prior treatments are not usually erasable. This is a particular problem for one-group designs of type 8 or 9.

But if, for some reason, these definitions are not understandable we will attempt a short synopsis at this point using common examples. Read their definition first and this example second, if needed.

1. History - It is not uncommon for experimenters in instructional research to forget that the textbook is a study material which suppliments any lecture or movie which may be the treatment and that some students possibly either in the treatment or control groups have access to the textbook. It is extremely difficult to control for outside studying in instructional research.

2. Maturation - Long term curriculum studies which observe students over time periods in excess of three years find great difficulty attempting to control the effects of maturation.

3. Testing - Providing students with an exam which gives some indication of the outcomes of instruction usually increases the IR scores on post test. In attitude studies the administration of a pretest usually depresses the intensity of the attitude change.

4. Instrumentation - It is common in studies where a number of ratings of teachers or students are made over a period of time that the raters become more proficient and therefore change as a measuring instrumenter. Frequent checks of the reliability of the rater and his accuracy in making the judgements should be planned for in the design.

5. Statistical regression - Theorists many times concern

themselves with a particular group of individuals, e.g. High IQ. Unfortunately in multiple testing of these extreme groups, their scores statistically regress toward the mean and wash out possible treatment effects. Any attempt at random selection of extreme groups, therefore should be carefully considered in the design.

6. Biased selection - Until recently it was not uncommon to find subjects being matched on a series of variables thought to be highly related to the treatment variable. Unfortunately, 1) all of these selection variables cannot be controlled, and 2) the loss of students who are unmatchable is large and unestimable.

7. Experimental mortality - Sickness and dropouts from classes tend to diminish the conclusiveness of the results of any experiment. One way to prepare for these contingencies is by having a large sample.

8. Interaction of Selection and Maturation - There are occasions when one wishes to use the entire sophomore class within the experiment, seeks to find a comparable group and the best he can locate is a junior class. The experimenter runs the risk of this particular source of invalidity since the juniors with their year of growth may in fact be different than the sophomores. This maturation may produce a gain that can easily be mistaken for the effect of the experimental treatment.

9. Interaction of Testing and Experimental Treatment - A pretest has the general effect of increasing scores on a cognitive test and decreasing scores on an attitude test. It can be easily seen that a pretest is a practice of the post-test. Also it provides the student with a set toward the learning material in the sense that he knows what is to be expected. In an attitude study the pretest also has "set" to introduce the student to the particular attitudes that the film or whatever the experimental treatment wishes to influence. The student is presumably much more conscious of the manipulation and less likely to change in the desired direction.

10. Interaction of Selection and Experimental Treatment - In many experiments students are asked to volunteer their services as subjects. This seemingly unimportant fact that the students volunteered may interact with the particular treatment to bias the results. Here again, this source of invalidity demonstrates the tremendous problem one has in selecting students by matching or some other way other than by random assignment.

11. Reactive Effects of Experimental Arrangements - It is extremely difficult to remove this source of invalidity since every student usually is aware of the fact that he is a member of an experiment. We all try for the "candid camera" effect but many times it is impossible to bring off.

12. Multiple Treatment Interference - In the more complex experiments we find that when one has multiple treatments he sometimes wishes to try each of these treatments on every student, thus encountering a host of possible errors, e.g., the effect of the first treatment upon the second and so on.

Now read Practice Exercise #1. As you read the proposal, answer for yourself the following four questions:

What are the experimental treatments?
What are the experimental units?
What is the method of assignment?
What design category does this project seem to fit?

You will note that these are the same four questions given above on your checklist. After you have answered each question you may compare your answer with those given below.

<u>Answers to Practice Exercise Number One.</u>

What are the experimental treatments?

<u>ANSWER</u>

Although the experimental treatments are given a heading, a closer inspection reveals them to be terribly complicated and multi-faceted. This is a typical case of too many eggs in one basket, and if certain effects aren't forthcoming there is no way of determining which of the number of different treatments has caused which effect.

What are the experimental units?

ANSWER

> At first reading it seems that four groups are going to be considered the experimental units, but further on we find that subjects or students will be assigned to each of these treatments.

What is the method of assignment?

ANSWER

> It is obvious both that some form of matching is hinted at but not explained clearly and herein lies a second fundamental flaw in this proposal.

What design category does this project seem to fit?

ANSWER

> In searching for R's, X's and O's, we find first that there is no random assignment but some form of matching. Second, all groups receive one kind of treatment. Thirdly, there are weekly observations on each student. Having more than one measure on each student, as you can well see, further complicates this already complex proposal. A designation of this study in one of the six categories would depend primarily on the use of the observations. If the first Monday-morning test were to be used as the pretest, then it is possible that they look like a group 2 design, where each of the experimental treatment groups is its own comparison.

If we quickly run through the sources of invalidity of this particular proposal we find that history is alluded to in the selection process but not really specified in the detail needed for this kind of study. Maturation is not of too much concern because the study duration is only one term. Each group or each individual in each group receives a test Monday morning therefore somewhat controlling for the effects of testing upon further tests. (Instrumentation) Having a single rater score each of the student responses especially with his own numerical scheme is wide open for criticism. If a rating scheme is used then a number of raters, possibly four or five should be used and their results correlated to determine their reliability and consistency over

time. The writer says nothing about selecting groups on the basis of extreme scores, although the selection criteria are not well spelled out in the first place, so the statistical regression cannot be considered a fault. The next source, number six, biases resulting from differential selection, is again an unknown but obviously a tremendous problem in this proposal. No mention is made for any attempt to control experimental mortality or student dropout or sickness. In small classes such as these, fifteen, it becomes crucial then if students stay away from class for a couple of days and lose their academic place . In the course of an entire semester there is no telling what effect this will have on the overall results. Again, when considering the interaction of selection and maturation the lack of specificity in detailing the selection techniques makes it impossible to make a judgement in this case. The last four sources of invalidity are determined primarily through interaction with the experimental treatment. Since the experimental treatments are not well defined then it becomes impossible to make judgments on these four as well.

One can tell in reading this study that it is a typical instructional situation where an innovative set of ideas is trying to be implemented into the present instructional system. This is exactly what we are attempting in developing better, more challenging college instruction. Yet the proposal is a perfect example of the tremendous complexity of college instruction and the difficulties in doing long, term course length, studies where so many intervening types of activities come between the student and the total effect upon him of the course. The best advice we can give to this proposal writer is to make much more specific his experimental treatments and attempt to assign some theoretical tie between them and the tasks that are being learned. With this theoretical tie made he can make some fairly specific hypothesis and tie the study down in terms of time and complexity to shorter periods, perhaps a day or a week where short term studies of the effects of these variables can be culled out. The writer has the making of these smaller studies when he considers the weekly testing. He might capitalize on these weekly tests in an attempt to show differential gain over a period of time rather than the one lump sum approach.

Now read Practice Exercise #2. As you read the proposal, answer for yourself the following four questions:

What are the experimental treatments?
What are the experimental units?
What is the method of assignment?
What design category does this project seem to fit?

You will note that these are the same four questions given above on your checklist. After you have answered each question you may compare your answer with those given below.

Answers to Practice Exercise Number 2.

What are the experimental treatments?

ANSWER

One of the treatments obviously is the using of the electronically modified piano. It is not stated what the subject or pupils will do with this piano but we can assume that he plays it and it plays back to him somehow. A much more difficult problem arises in the notion of conventional instruction. Entirely too many studies rely upon this gimmick to back away from explaining what actually does happen to these students. It is relatively obvious that the writer doesn't know and each student probably receives some different instruction. The writer has tried to camouflauge this by saying that it is an instructional method widely used throughout the country.

What are the experimental units?

ANSWER

It is obvious that subjects or students are the experimental unit. But it is not so obvious that members of one group receive the same treatment as members of the other. For example, each student seems to be studying independently of the other sutdents. Therefore one may study considerably longer while another student may be studying one key and not another key. Is it correct then to assume that each of these groups of students are receiving "the" experimental treatment?

What is the method of assigrment?

ANSWER

Here the writer intimates that students will select different sections of the course by chance by saying that "it is reasonable to assume that no factors other than chance are operating in the assignment of individuals to groups." Unfortunately, that is not an easy assumption to accept for the simple time allotment to each section many times determines the type of student entering each

Continued

ANSWER (Continued)

of the sections. Again, anything less than random assignment lays the writer open to much criticism. The writer's reference to the further expectation that rotation of treatments will compensate for any differences in assignment to groups only confuses the matter more. If a student increased, let's say 10 points in the first group, what is the effect of that increase on the second is an experimental question, not one that easily eliminates itself.

What design category does this project seem to fit?

ANSWER

Again, this study, like the previous one is categorized somewhere between three and six. Since there was no true randomization and only one observation made after the treatment we have to put it somewhere between these two.

Let us run down the list of sources of invalidity. As was mentioned earlier, what happens between that first test and the last is anybody's guess. When the treatment is unspecified, history effects also are under question. As in the last study, maturation effects are fairly well controlled since each of the groups mature together. The question of testing effects becomes somewhat complicated when one considers all the testing that is done during the course itself. The instrumentation problem is handled with sophistication in this study with three raters, each of which does not know the identity of the student. Indices of inter-rater reliability are to be determined and reported. Statistical regression effects don't seem to be of concern because extreme groups are not in use. Again, the problem of selection is left somewhat in doubt because of the gross assumption that students electing courses will be the same as a random assignment. With only 20 students per class the experimenter is flirting with danger when not considering the mortality problem. The interaction of selection and maturation remains a problem as the selection procedures are somewhat under question. Since the pretest was simply an initial proficiency test it might be argued that there would be no interaction with the treatments although one cannot really say. The selection of experimental treatment effects again are based upon the weak assumption underlying the selection procedure. The reactive effects of the experimental arrangements is an extremely problematical area in this research. Those using the electronic keyboard certainly know that they are being watched, so to speak, and those not using the electronic keyboard are envious of those

who are. How one controls for these reactive effects is usually through some form of counterbalancing, giving A then B for one and B then A for another. As in the first study, breaking the experiment up in much smaller parts and giving one student a try at the piano with certain conditions, and another student at another time, each student is affected by each treatment. But in doing that one falls into the obvious pitfall of the last source of invalidity; mainly, the multiple treatment interference. It is entirely possible that the student who begins the electronic keyboard may become very depressed with the outmoded conventional technique.

In this experiment as the one that came before it, the experimental treatments were not consistent over individuals in the experimental and control groups. Therefore, it might be well to proceed on an individual basis recording the rate of learning of each individual and perhaps through counterbalancing of treatment and conventional or electronic and nonelectronic instruction one can determine the effectiveness of one over the other. Again a caution must be made in making the task too large, that is, considering an entire semester's course of study. It would be much more appealing if one would consider each of the tasks, perhaps scales left hand, right hand, or any way of breaking down the total task into smaller units and looking at the effect of the keyboard on the smaller units of study. By using the smaller units and possibly interchanging students from one instructional method to another, it would then become possible to randomly assign students to treatments rather than have to be constrained by the administrative scheduling of the courses. Again, it seems that this type of study could have been much better carried out in a one to one basis or a laboratory approach where one student or each student would be watched carefully over a period of time to determine the influence of either of the two methods.

With the time remaining, read Class Exercise #1 (Appendix C). Answer for yourself the same five questions for this proposal and be prepared to discuss the proposal during class.

Appendix A

Practice Exercise #1

A Study of the Use of Motion Pictures In Foreign Language Instruction

I. <u>Problem</u>:

Recent surveys reveal extensive numbers of foreign language laboratories in use in the schools, colleges, and universities of the United States (Social Science Research, Inc., 1963). That most of these laboratories are being used ineffectively is well documented (T.C., Columbia Univ., 1963). The purpose of this proposed research is to investigate a new method for increasing the effectiveness of language laboratories. The method to be investigated makes use of motion pictures, and was designed as a result of accidentally obtained pilot data which indicated the feasibility of using motion pictures to enhance language laboratory effectiveness. The accidental finding resulting from delayed receipt of materials suggests that motion pictures will enhance the effectiveness of the language laboratory to a significant degree if they are used following the language laboratory experience but will impe' learning if used in conjunction with the laboratory experience. There are good theoretical bases for the position.

II. <u>Related Research</u>:

In the investigation comparisons will be made of the performances of students following three types of learning experiences. Students in a course in Modern Spanish will be studied. Selection of Modern Spanish was seen as unique because (1) the content has been agreed upon by experts in the field and (2) well developed materials are available for use. The use of the visual materials proposed in this study is quite diverse from that reported by Creore and Hanzeli (1960). The method reported by Creore and Hanzeli will constitute one of the controls in the proposed experiment. A more complex design than that used by Creore and Hanzeli will provide a more adequate test of hypotheses, pertaining to the use of films. The proposed use of visual materials also differs from that employed by Asher (1961) in which subjects learned a Spanish vocabulary list using one type of stimuli (visual or auditory) and relearned the list using the

other stimulus mode. Because of the divergent use of the visual stimuli by Asher, no control of his techniques will be made in this study.

Extensive research in the acquisition of foreign language skill has resulted in numerous guidelines for foreign language instruction (c.f. Modern Language Association, 1961; Lane, 1961). Little evidence is available, however, regarding the appropriate use of motion pictures in the foreign language classroom or laboratory.

There have been numerous uses of motion pictures in instruction. Films have been used to introduce units of study, to present information, to motivate thought and activity, to modify concepts and attitudes, to integrate a series of facts, ad infinitum. All of these uses might be readily applied to foreign language instruction. It is the purpose of this study to investigate three possible uses of a set of motion pictures in a beginning Spanish course for college students.

The Spanish Course. Following suggestions of the Modern Language Association of America (MLA), the faculty of Southern Oregon College (SOC) now offers an elementary course in Modern Spanish. The primary objectives of the course are to teach students to speak and to understand conversational Spanish. Behaviorally the objectives might be stated as follows:

Following the course of instruction the student:

1. Responds with correct grammar, tone inflection, and accent, in Spanish, to a series of Spanish patterns, expressions and questions.

2. Converses in Spanish with another individual about any of a number of common persons, places, or things including ideas.

In other words, the course is intended to prepare a student to the point that he could express himself adequately and without language related embarrassment in the Spanish speaking community such as Mexico City. It is not the primary purpose of the course to train students to read or write Spanish. This is consistent with the suggestions of MLA that students first develop speaking and comprehension fluency and later reading and writing proficiency.

A set of materials, including a text (MLA, 1960), magnetic tapes, and a series of correlated motion picture films, which were prepared under the auspices of MLA will be used in the Spanish Course at SOC. The text is divided into 30 units, of which 24 are developed around a conversation (dialogue) between two or more persons about the same topical episode. Each dialogue has associated with it a tape recording of the dialogue (conversation), for use in a language laboratory, and a motion picture film depicting the dialogue. MLA suggested that the dialogue tapes and films be used for introductory experiences in the classroom and that the laboratory tapes be used individually by the students for practice.

Upon initiation of the course at SOC in September, 1962, the films were not available. When the films arrived in October, 1962, the instructor began to use them in two ways--(1) as an introductory experience for future dialogues and (2) as a review experience for dialogues which had been previously completed with the use of the dialogue tapes but without the use of the films to accompany the dialogues. Informal evaluation of student learning indicated that use of films seemed to result in differential amounts of learning and retention. The evaluation further suggested that learning was more complete for those dialogues in which the film was used for review than for those in which the films were used as an introductory experience. It is felt that a thorough investigation of these informal findings is merited.

<u>Theoretical Bases</u>. The use of motion picture films as a review or culminating experience is consistent with behavioral theories which state that before a response can be used in appropriate context it must first be in the repertoire of responses of the individual. This suggests that the development of a response repertoire in Spanish by the learner precede his attempts to use the responses in a meaningful fashion. Lane (1961) stated that the development of the response repertoire results from the application of principles of operant conditioning in the learning process. The activities in the classroom and language laboratory are roughly analogous to this type of conditioning although Lane and other Skinnerian psychologists would probably find the initial responses required in typical Spanish classrooms and laboratories more complex than they would wish. In other words, after the responses have been acquired, they will then be associated with a meaningful situation. This arrangement, however, is only one approach that will be investigated. In the proposed study the filmed versions of the dialogues will be used to provide meaningful experiences containing a large number of visual stimuli, to any one of which the verbal response may be associated. Tape recorded versions of the dialogues will also be used to provide

this experience; however, it is hypothesized that appropriate use of the visual stimuli presented in the motion pictures will enhance the meaningfulness and the retention of the total learning experience.

A second experimental use of motion pictures will employ them as an introductory experience. This application has as its theoretical base a Gestalt explanation of the learning process (Hilgard, 1956). This orientation suggests that learning is more efficient and lasting when it occurs within a field which is structured either by or for the learner. The total or whole experience is assumed to be a more meaningful one for the learner and one to which his responses may be related readily. This, apparently, was the basis used by Creore and Hanzeli in their experiment. Unique "control" conditions will be provided to help clarify the meaning of differences in learning which are expected.

III. Objectives:

The principal objective of the proposed study is to provide evidence regarding procedures to be followed in enhancing the use of language laboratories.

IV. Procedures:

Four groups of 15 students will participate in the study. Each group will receive instruction in a modern Spanish course under different treatments. The treatments are described below. Assignment of subjects to the experimental treatments will be done in such a way that the four groups do not differ in academic ability, achievement level, or previous experience with the criterion foreign language, namely Spanish.

Experimental Treatments. The four treatments are shown in Figure 1. The treatments are so designed that amounts of time spent by learners in the classroom and laboratory are the same regardless of the treatment. Study from text by the subject will not be controlled, however, each subject will be required to record and report the amount of time spent in this way. The experiment will be conducted during the first academic quarter of the course which extends over three quarters. In this way greater control over previous experience may be maintained. Each group will meet in class on Mondays and Fridays for one hour. On Mondays each group will be tested on the previous week's assignment, and will receive an orientation to the week's activities. Orientations will differ by treatments. Fridays will be devoted to differential review experiences which will be described in detail below.

All students will be required to spend five hours in the language laboratory between Monday and Friday, and must spend at least one hour but no more than two hours each day in the laboratory. The laboratory practice will be divided into two phases. A basic phase (D) will consist of practice on the actual dialogue to be mastered. Once this phase is perfected, each student will practice on an enrichment phase (E) which provides practice with variations of the content in the basic dialogue. Experience indicates that on the average each student will practice approximately two to three hours in the dialogue phase and two to three hours on the enrichment phase. Attendance in the laboratory will be required and will be limited to five hours per week.

The experimental design, thus, comprises four groups of students who will receive five hours of individual training in a language laboratory and two hours of class experience each week. The groups will be differentiated only in terms of differential experiences they receive in the classroom setting. The experiment will continue for one academic quarter.

Since the laboratory experiences for each treatment group do not differ, the description of treatments below will be restricted to the experiences received by each group in the classroom.

Treatment A. On Monday of each week, the subjects in this group will be tested over the training received during the previous week. This will constitute about 15 minutes of the class period. Following the test period, the subjects will be introduced to the dialogue to be studied during the week. Introduction experiences for this group will consist of listening to several repetitions of a tape recording of the dialogue. After each repetition the instructor will answer any questions and will identify "trouble spots" for the class. On Friday the group will meet with the instructor and again listen to several repetitions of the same tape recordings and discuss the dialogues and the enrichment tapes in the context of the history and literature of Spanish-speaking countries. This treatment represents typical approaches to language instruction now in use.

Treatment B. The experiences on Monday of each week are the same for this group as they are for subjects under Treatment A. On Friday this group will meet with the instructor and will listen to and view a motion picture film of the dialog studied during the week. The films will be repeated the same number of times as the tapes with duscussion organized in the same way as for Treatment A.

This treatment represents an application of the operant conditioning model discussed previously, and on the basis of the pilot data described above, this treatment is hypothesized to result in superior performance (criterion discussed below) by the students.

Treatment C. The experiences presented to subjects in this group will be the same as those for Treatment A except that under the treatment subjects will receive the introductory experience on Monday using the film version of the dialogues. This treatment represents an application of the Gestalt orientation and is similar to that recommended by MLA.

Treatment D. Experiences presented to subjects in this group will be the same as for those in Treatment B except that under this treatment subjects will receive the introductory experience on Monday using the film version of the dialogues. This treatment represents essentially the same treatment used by Creore and Hanzeli and is typical of one use of visual materials in language instruction.

Figure 1. The experimental Treatments

Day / Condition	Monday	Tuesday	Wednesday	Thursday	Friday
A	Test/tape	D/E*	D/E	D/E	Tape
B	Test/tape	D/E	D/E	D/E	Film
C	Test/film	D/E	D/E	D/E	Tape
D	Test/film	D/E	D/E	D/E	Film

* D represents individual practice with the dialogue tapes

E represents individual practice with the enrichment tapes

Evaluation Procedures. Students will be evaluated during the first portion of each Monday's class during the quarter. Criterion instruments will be administered individually and will consist of verbal responses to a set of standardized stimulus expressions in Spanish to which the students must respond in Spanish. The stimuli will be presented by the instructor and responses will be tape recorded for evaluation. Separate test stimuli will be prepared for each dialogue and will be based on the speech patterns developed in the dialogue. The instructor has had extensive experience in rating student responses and has developed a numerical scheme for recording and scoring errors made in fixed test conversations, student response, grammatical correctness, and accent. Numerical scores will be assigned by the instructor.

Analysis. It is hypothesized that use of the films as introductory experiences will have a tendency to inhibit the learning of the foreign language through the introduction of a negative set resulting from the impact of many irrelevant stimuli. However, the use of films after the verbal behavior (depicted in the films) has been learned will significantly enhance retention and performance because a much larger number of stimuli will become associated with and will elicit desired responses. Null hypotheses of differences between performance scores will be tested using analysis of variance techniques.

While the number of subjects in each cell is relatively small (N=15), relatively large amounts of instruction will be provided each subject, making it economically feasible to provide such instruction if the experiment indicates a preferred instructional method.

Approximate Time Schedule. Development and refinement of the measurement instruments, laboratory tapes, and course materials will be accomplished during the six month period from January 1, to June 30, 1964. The experiment will be conducted during the fall academic quarter. Analysis of the data and completion of the final report will require two months beginning January 1, 1965. The project will not be active during the months of July and August, 1964.

Expected End Product. The results of this research will shed new light on the appropriate use of visual materials in conjunction with foreign language laboratory experiences. It is expected that the test materials developed for the project will be suitable for use in a variety of Spanish teaching situations.

Publication Plans. The results of this research will be offered for publication in professional journals such as the Audio-Visual Communication Review.

Appendix B

Practice Exercise #2

I. Problem

In American education the expectation is virtually universal that the adequately trained elementary school teacher will have at least some minimum proficiency in keyboard (piano and organ) instrument skills. Music instruction, particularly in the primary and intermediate grades, is usually closely integrated in the instructional program, and classroom teachers are expected to be relatively self-sufficient in supplying music experiences for their students. Consequently, certification requirements and degree programs for elementary teachers commonly require either specific training or demonstration of proficiency in keyboard instrument skills.

This requirement presents music educators in elementary teacher education institutions with unique problems. The continuous and increasing demands for well-trained elementary school teachers, and the increasing number of students choosing this profession, exerts pressures for increasing class size and student-teacher ratios. However, the nature of classes requiring performance dictates that they be small and students have opportunity for individual practice and tutoring. It is generally accepted by music educators that students must receive frequent feedback regarding the adequacy of their performance if they are to progress rapidly. Thus we have an enigma with regard to efficiency. Small classes and low pupil-teacher ratios are inefficient in terms of utilization of staff and facilities, but large classes and high pupil-teacher ratios are inefficient in terms of student achievement.

The question of efficiency can be pursued further. Considerable specialized training and ability are expected of a music instructor, but teaching fundamental keyboard skills to naive students dictates that he listen to students perform and tell them how they are progressing. Much of his time and effort is devoted to rudimentary and repetitious instruction, not markedly different from that which would be given a young child in private instruction. If there is no alternative, the situation can be accepted as a necessary evil, no matter how regrettable it is that much of the instructor's capabilities are not exercised. But if an alternative exists, then it appears obvious that the feedback-giving role is an inefficient use of instructor's time.

The need to acquire keyboard instrument skills presents the student with certain problems, too. The elementary education teacher candidate who has had no previous musical training is the typical case, not the exception. As with dancing, swimming, and many other skills demonstrated in public performance, the most acceptable time for learning has passed, and public demonstration of ineptitude is humiliating. Yet the nature of the learning task requires some active performance, however inept, from him, and that he be given feedback regarding the quality of his performance. Private tutoring for all students is simply not feasible. Practice without feedback is inefficient because errors are practiced along with correct performance. And when performance and feedback take place in a class setting, the feedback is not merely informative, but often threatening to the student.

Briefly stated, then, problems to which this proposed project are addressed are threefold:

1) How can more adequate tutorial feedback be provided to beginning piano students in elementary teacher training programs?

2) How can instructor time required for rudimentary and repetitious phases of piano instruction be reduced?

3) How can the adverse effects of public practice and public feedback on students learning basic piano skills be minimized?

One stage in the process of acquiring basic piano skills where problems mentioned above seem to be concentrated is the point where students learn the basic chord patterns in various keys. Learning these patterns is essentially a rote process. After seeing the correct patterns demonstrated on a piano keyboard students go through a repetitious process of individual practice, performance and appropriate correction by the instructor.

Oregon College of Education and the Teaching Research Division of the Oregon State System of Higher Education are presently engaged in a pilot study which involves the development and testing of an automatic training device to be used in teaching this skill. The music staff believes that this learning task is particularly well-suited to self-instructional procedures. The educational objective may be classified, essentially, as the acquisition of a psychomotor skill.

Correct performance is readily distinguished from incorrect; either the correct combination of keys is struck, or it is not. The striking of the keys is, in a sense, a switching process. The piano mechanism translates a finger movement into a musical note; a simultaneous combination of movements becomes a correct and pleasing chord, and an incorrect combination, a discord. Modification of a piano to accomplish automatic switching through electronic means is a relatively simple procedure that need not interfere markedly with its more conventional function.

Such a device has been constructed. A piano has been modified to accomplish switching, as described above. The striking of a combination of keys will be converted electronically into input for computer-type logic circuitry capable of discriminating the correct pattern from all others and reacting with an output that can be translated into meaningful feedback to the student as to the correctness of the pattern played.

Capabilities of this device include: indication to the student of the chord pattern to be played, visual and/or auditory feedback to the student of the correctness of his performance at each step of a chord progression, recycling to the beginning as soon as an error is made, automatic advance to a new chord progression as soon as a correct progression has been completed, recycling of a given progression for a predetermined number of times (e.g., six consecutive correct performances before automatic advance), or manual control by the student of the chord pattern and progression to be practiced.

The use of this device may free the instructor to a large extent from that portion of his instruction that involves repetitiously evaluating and giving feedback on student performance. The student may be enabled to practice privately, with continuous, accurate, unthreatening automatic feedback regarding the quality of his performance. Outcomes of the pilot study now being completed include a prototype of the device, schematic drawings and descriptive reports that will enable reproduction of the device, determination and improvement of the validity and reliability of instruments to measure performance of the instructional objectives and affective reactions to its use, and possible but presently unanticipated changes in the design of the device. The purpose of the proposed research is to assess the extent to which use of this device may alleviate the three-fold problems described above.

II. Related Readings and Research

The review of the literature for this study is confined to two separate areas of research: (1) psychological studies which involve the use of feedback and (2) music educational research.

Feedback. Various studies of knowledge of performance, (McPherson, Dees and Grindley, 1948-49; Michael and Maccoby, 1953, Ammons, 1956) have shown that students improve in their performance when they are given knowledge of results. Michael and Maccoby concluded that "the most important factor in influencing the amount of learning in this experiment was the provision of knowledge of correct response (KCR)." Wolfle (1951) stated that "laboratory studies are unequivocal in emphasizing the importance of giving a subject s specific and as immediate information as possible concerning the outcome of his efforts." (p. 1267)

In 1954 Skinner suggested that laboratory research on behavior had a direct bearing on the teaching process. He felt that application of research findings could be effectively brought about by using a mechanical device. This device was to supply the student immediately with knowledge of his correct response. In 1961 he stated that "exploratory research in schools and colleges indicates that what is now taught by teacher, textbook, lecture or film can be taught in half the time with half the effort by machine."

Music Education Research. The reports of several studies of late (Barnes Buchanan, Carleen, Woelflin) point to the desperate problem faced by music educators in colleges which train elementary teachers. Plainly stated, the regular classroom teachers are not trained adequately to do an effective job of teaching music. Buchanan (1964) stated "The ability to play a piano is an asset to a regular classroom teacher, and often times it is the determining factor in being selected for the job... Something must and can be done to compensate this...(training)... deficiency. Just as in science, mathematics, languages and other areas of learning, music education must streamline and revise its methods and procedures of instruction. Certainly this is true of piano teaching. The purpose of such a revision is to more effectively and more economically, from the standpoint of time, prepare the prospective teacher to meet his pianistic needs."

Carlsen (1962) used programed instruction to develop melodic dictation ability and concluded, "The results of the experiment clearly indicated the value of the (method)... The potential of programed instruction appears great...to release the teacher for

tasks which only the teacher can do." Woelflin (1963) experimented with instrument instruction by program machine to free the teacher from tasks which amount to supplying feedback. Barnes (1963) introduced programed instruction over factual information as such not only "could save many hours of instructor time and student time, but the use of the programed book could permit the teacher at the outset of the course to assume a specific level of competence on the part of every student in the class."

A call for help in the matter of research itself was issued by Petzold at the Music Educators National Conference, 1963. He commented that "during the period 1952-62, 70% of the music education dissertations listed in Dissertation Abstracts were the relatively uncomplicated and highly popular survey studies and 30% were basic or action type. Often the survey was made in desperation and did nothing but help perpetuate the commonplace in music education research." He stressed the need of professional team work, i.e. "the teacher must: (1) be relieved of a portion of his teaching load in order to have sufficient time to carry on investigative activities designed to improve the program; (2) be given assistance in planning projects, and consultative services by trained researchers should be available throughout the course of the projects. Such cooperation...will... result in substantial gains toward improvement of programs and instructional procedures."

III. Objectives

The proposed research seeks to determine:

1) Will use of the training device described eliminate or significantly reduce the amount of time required of the instructor for tutorial or feedback-giving activity?

2) Will students using the device demonstrate at least as much increase in proficiency in playing the selected chord patterns on completion as those taught in a described conventional manner?

3) Will students using the device require equal or less practice time to reach a criterion of performance than those taught in a conventional manner?

4) Will student attitudes, as measured by a Thurstone scale, indicate significantly more positive affect toward learning of the skill in question by students using the device, than by students receiving "conventional" instruction?

5) If differences in affect are noted, are they due to novelty of the device, its effectiveness, or the privateness of instruction?

6) After working with the device, will teachers indicate a preference for its continued use, or to return to conventional instruction?

IV. Procedure

a. General design and selection of sample.

Two groups of subjects will be used in the study, an experimentally taught group and a conventionally taught group. Subjects will be selected from students enrolled in the Music Fundamentals classes at Oregon College of Education, Fall term, 1965, and Winter term, 1966. Classes of approximately twenty students each will be used Fall term, and two other classes of similar size will be used Winter term. Since two instructors teach the course each term, they will be rotated in the experiment so that the instructor of the experimental group the first term will have the conventional group the second. Rotation for time of day will be accomplished similarly. Subjects will be pretested to determine initial level of performance. Prior experience indicates that students with initial proficiency are relatively rare, but all such students will be eliminated from the experiment. Since students in these classes are in the same educational program and will not be informed of instructor or treatment until after enrollment, it is reasonable to assume that no factors other than chance are operating in assignment of individuals to groups, with the further expectation that rotation of treatments will compensate for any differences in assignment to groups that may exist.

b. Experimental treatment

In both experimental and conventional groups the instructional objectives will be the playing of four-step chord progressions in eleven keys, namely G, D, F, B^{b}, A, E, E^{b}, A^{b}, C_1, C_2, C_3, with acceptable tempo and regularity. Subjects will be informed of the objectives at the outset of the study.

All subjects will receive similar instruction in music fundamentals prior to introduction of the experimental treatment. Such instruction will include basic elements of notation and a theoretical explanation of chord building. Sullivan notebooks will be used by all subjects. Differences between experimental and conventional groups will begin immediately following demonstration of the basic chord progression.

From this point to the post-test, subjects in the groups using the electronically modified piano will not perform in class or receive feedback from the instructor. All practice and feedback will be accomplished through use of the device. In the regular class meetings for these students, other areas of musicality, such as music literature, will be presented.

The conventional group, during this same time, will experience an instructional method widely used throughout the country. This will consist of each student in turn performing certain chord progressions, receiving feedback from the instructor regarding his accuracy, and witnessing the same process enacted with other students.

All subjects, experimental and control, will be directed to request a post-test when they have completed five consecutive correct progressions in each of the eleven required keys at a single sitting. In the case of the students using the electronically modified piano, the device will automatically advance when the desired number of correct repetitions has been accomplished, so his readiness to take the test will be indicated to him automatically. Control group subjects will use a checksheet to record their consecutive correct progressions to determine completion of their training period. The post-test will be administered within 36 hours of said completion, and will consist of the students' actual performance on eleven progressions, each of which will be tape recorded.

c. Data and Instrumentation

1) Time records.

a) Instructors of both experimental and conventional groups will keep accurate records of actual time spent in a tutorial feedback-giving relationship with students. While it is expected that the instructor of the group using the electronically modified piano will spend no time in this manner, unanticipated events may necessitate some such activity.

b) Practice time cards will be distributed to all students and also will be placed on all school pianos. Subjects will record all practice time, whether spent on school or home pianos or in performance in class. Total time spent by each student will be computed.

2) Post-test scores.

The tape recorded post-tests will be evaluated by three competent music instructors, who will not know the identity of the subject whose test is being evaluated, or whether he was in the experimental or conventional group. Indices of interrater reliability will be determined and reported. The score assigned to each subject's performance will be the average of the three scores given by the evaluators.

Post-test scores will be weighted 60% on accuracy, 25% on tempo, and 15% on regularity (steadiness). The accuracy score will be determined by scoring each of the four steps in each of the eleven progressions as either correct or incorrect. Tempo will be scored on a five-point scale, ranging from one point for less than 40 chords per minute to five points for 96 or more chords per minute. Steadiness, or regularity, refers to the fluidity or lack of hesitation demonstrated in playing the chords. This also will be rated on a five-point scale.

3) Thurstone scale.

A Thurstone-type scale will be constructed to measure the affective reactions of the experimental and conventional group subjects to the type of instruction in chording that they received. In addition, anonymous, open-ended written statements will be collected from the students, to discover unanticipated student responses to the treatments.

4) Instructor reactions.

Since the members of the music department have already expressed enthusiastic support of the project, it does not appear that a systematic comparison of their reactions to the respective treatments would be appropriate. However, a narrative report on their experiences will be made, in

which they will indicate whether, after using the training device, they would prefer to continue using it or to return to conventional instruction, how they would like to see the device or its application modified, and whether they would like to see an extension of this training principle into other areas of their course.

d. Analysis of the data

Significance of the difference between the experimental and control groups in instructor feedback time, student practice time, student post-test scores, and affective reactions of students to the respective treatments, will be determined by t-tests.

e. Approximate time schedule

July 1 - September 30, 1965 - Modification of two additional pianos; construction of measurement instruments; orientation of staff.

October 1, 1965 - March 31, 1966 - Experimental treatment; data collection.

April 1 - June 30, 1966 - Data processing; writing of final report.

Outcomes of the proposed research effort will include:

1) A research report providing evidence on the questions asked above.

2) The existence of three use-tested devices for instruction in chording skills, available for demonstration to music educators and research specialists.

3) Schematic diagrams and descriptive literature to allow replication of the training device.

4) Information regarding possible modification of the device or its utilization.

Appendix C

Class Exercise #1

Procedure

General Method and Research Strategy

The general plan of research is to compare changes in performance of five, randomly assigned groups of inservice teachers on criterion measures developed by Schalock, Beaird, Simmons, (see Appendix) for recording teacher behavior. The five groups represent the presence or absence of three conditions: (1) a workshop experience dealing with observation, self-evaluation and goal setting; (2) self-observation via video recording; and (3) overt self-evaluation via video-recording. The design permits the investigation of effects of self-evaluational procedures using video recording, workshop experience, and overt self-evaluation procedures on behavioral change.

Sample Plan. Sixty volunteer teachers from the Corvallis Public School System will take part in this study. During the first year (1966-67) thirty teachers will participate with twenty teachers in Group I and ten teachers in Group II (see page 12). The second year (1967-1968) thirty teachers will participate as follows: ten in Group III, ten in Group IV, and ten in Group V (see page 12).

The use of volunteers in this study deserves consideration at this point because important differences may exist between teachers who volunteer for self-improvement experiences, such as those provided by this study, and those who do not volunteer or must be directed to take part by some superior authority.

Following the pilot project (outlined in Related Research above) all teachers who had seen the demonstrations were willing to take part in a further study of self-evaluation via video recording. This willingness to volunteer is considered to be a major factor in the potential success of this study. Because the initiative for this study comes from the Professional Improvement Committee, Corvallis Education Association of the Corvallis Public Schools, i.e., the teachers themselves, (the Principal Investigator is Chairman of this committee) the variance between volunteer teachers and those required to take part in such a study cannot be examined legitimately. If self-evaluation through video recording is found to be a successful method of changing

teacher behavior, further studies will undoubtedly examine this variable.

Even though recording procedures developed by Schalock, et al. have been validated for the lower grades, because the home room situation both emphasizes the interpersonal communication processes observed and quantified and provides a more advantageous recording situation, volunteer teachers in the third through eighth grades will be used in thi sample. If sufficient teachers volunteer, this range in grade level will be collapsed downward to provide as much homogeneity on this variable as possible. Specifically, the Corvallis public schools employ approximately 25 teachers at each elementary and junior high school grade level. Of these 25 teachers at each grade level approximately 20 have had more than three years' teaching experience, but fewer than 10 years' experience. Since 60 teachers will participate in the study on a voluntary basis, it will be necessary to sample from a minimum of three grade levels. Attempts will be made to restrict the sample to third, fourth, and fifth grade teachers with the bulk of the sample being drawn from the fourth grade and the other two grades being sampled equally.

Treatment and procedure. The five groups of teachers taking part in this study will experience one or more of the following conditions:

1. Condition W (Workshop). A two-day workshop including training in video recording observation, using pilot project films, and classroom simulation films (Kersh, 1963); evaluation procedures, and the setting of behavioral objectives. Emphasis throughout the workshop will be on the development of personal, overt standards of evaluation with a view toward self-evaluation. This workshop will be conducted by the Principal Investigator, Research Director, and resource staff from the Teaching Research Division of the Oregon State System of Higher Education. The workshop will use observational procedures developed by Schalock, Beaird, and Simmons (1964), classroom simulation films and techniques developed by Kersh (1963), and a slide tape presentation designed to train persons to identify and state behavioral objectives developed by Paulson (1965). All the above named researchers are currently in the Teaching Research Division and can be used as resource persons at the workshop. The workshop will be presented at the beginning of each year of the study.

2. Condition SO (Self-Observation). The viewing of a playback of the video recording by the teacher, alone in the room, for the purpose of self-observation

3. Condition SE (Overt Self-Evaluation). Although it can be assumed that some sort of self-evaluative procedures are followed by all individuals confronted with a record of their own performance, whether or not the teacher writes down his criticisms of himself and makes written statements of his behavioral goals in the form demonstrated at the workshop is a controllable factor and differentiates this condition from condition SO.

Treatment Groups. During each of the two years of the study, thirty teachers will be randomly assigned to treatment groups to experience one or more of the conditions listed above. During the first year, two of the five treatment groups will be used to test the major objective which is: self-observation and self-evaluation will result in an over-all improvement of teacher effectiveness as defined and evidenced by her management behaviors and interpersonal orientation.

Group One. Twenty teachers will attend the workshop and participate in the self-observation and self-evaluation procedures. Because this group will experience all experimental conditions (workshop, self-observation, and self-evaluation), it is desirable to expose as many to the treatment as possible; therefore twice as many subjects will be utilized in this group as in other experimental and control groups. This group will have three recordings of their classroom behavior made and will view the first two. All three recordings will form data for analysis (W:SO:SE, N=20).

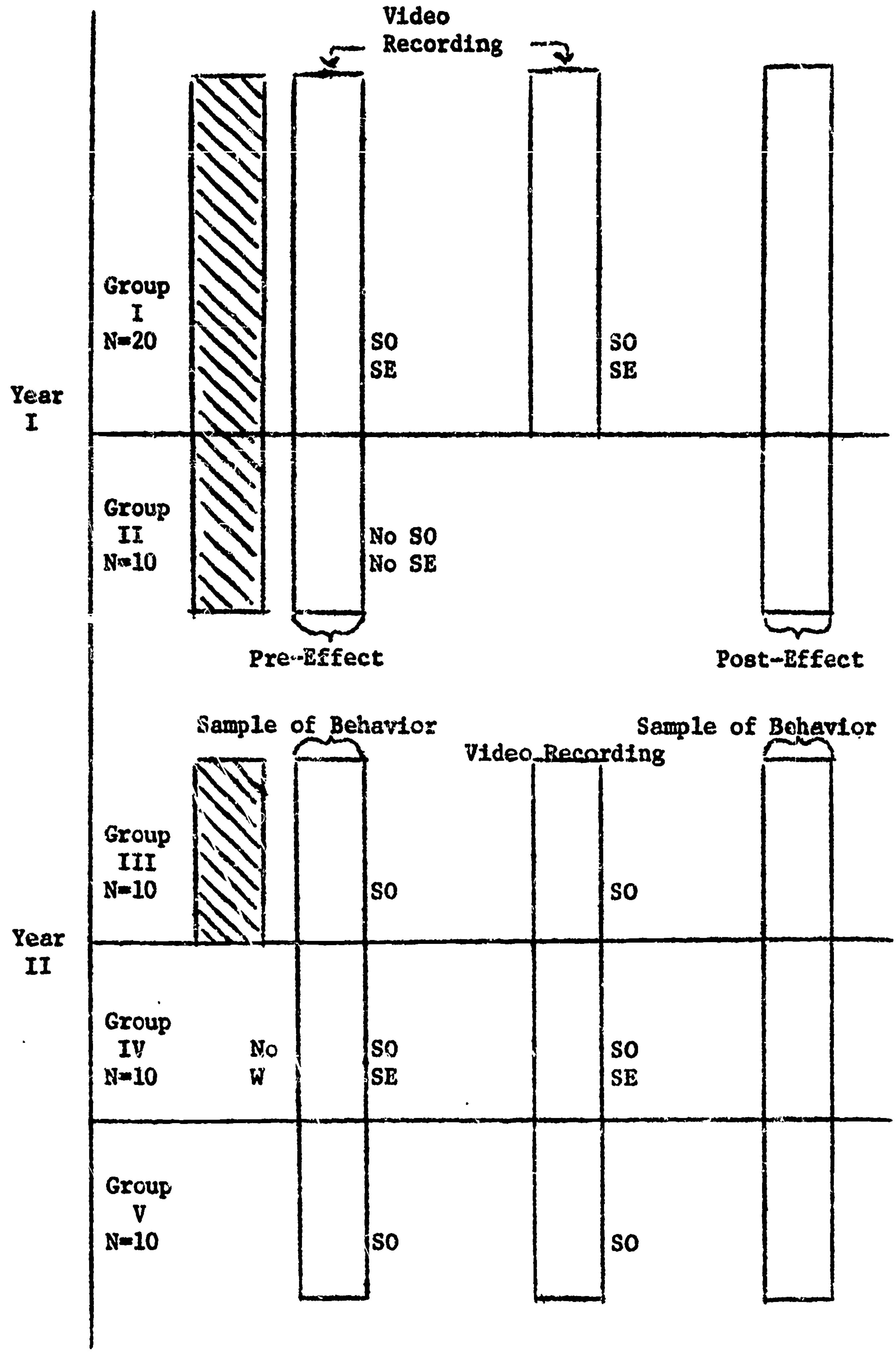
Group Two. A control group of ten teachers will attend the workshop and have pre- and post-experiment video recordings made of their classroom behavior. They will not participate in the self-observation or self-evaluation procedures (W:no-So:no-SE, N=10).

During the second year three alternate conditions (as listed below) of the experimental group will be tested in an attempt to evaluate and differentiate the effects of the workshop and overt evaluation processes. Groups three, four, and five will all be recorded three times and will observe their own behavior after the first two recordings. All three recordings will form data for analysis.

Group Three. Ten teachers will attend the second workshop and participate in the self-observation process; they will not go through the overt self-evaluation procedures (W:SO:no-SE, N=10).

Group Four. Ten teachers will not attend the workshop but will participate in the self-observation and self-evaluation processes using the forms provided for self-criticism and overt goal setting as best they can (no-W:SO:SE, N=10).

Group Five. Ten teachers will not attend the workshop, will participate in the self-observation process and will not go through the self-evaluation procedures (no-W:SO:no-SE, N=10).


Video
Recording
Group
I
N=20
SO
SE
SO
SE
Year
I
Group
II
N=10
No SO
No SE
Pre-Effect
Post-Effect
Sample of Behavior
Sample of Behavior
Video Recording
Group
III
N=10
SO
SO
Year
II
Group
IV
N=10
No
W
SO
SE
SO
SE
Group
V
N=10
SO
SO

VII - ANALYSIS I

Evaluation Sheet

We need to know how well the ideas and issues in this manual are communicated to you. You are the test audience for this material. To remove or strengthen the weak spots, to retain or improve the strong ones (if any), an account of your learning experience as you read these sections is crucial. Use this sheet as you study and jot down your reactions. Don't be concerned with typographical errors. We're concerned with how the message is coming through.

SECTION #________ TITLE______________________________

Give topic, paragraph or sentence, page	Importance Rate 1 2 3 4 5 Low - High	Clarity Rate 1 2 3 4 5 Low - High	Suggested Improvements

Over-all Rating of the Section: Use a 5-point Scale

Important

1	2	3	4	5
Very Low				Very High

(Circle one)

Understandable

1	2	3	4	5
Very Low				Very High

(Circle one)

PLEASE PUT ANY ADDITIONAL COMMENTS ON BACK

SECTION VII
Data Analysis I

Table of Contents

SECTION VII

Data Analysis I

James H. Beaird

In determining the appropriate analysis technique one must consider the following factors:

1. What is my research question?

2. What is the nature of my data?

3. How many samples do I have?

4. Are my samples related or independent?

Three types of research questions will be considered: (1) Experimental or Comparative, (2) Relational, and (3) Descriptive. For each type of question one of several techniques may be used for analysis depending on other factors, e.g. the level of measurement available, the number of samples, the independence of the samples. Let's consider each of these in turn.

Levels of Measurement

Four levels of measurement may be identified. These levels are defined in terms of the scaling characteristics of quantifiable descriptors (numbers) each provides. The four levels are: nominal, ordinal, interval and ratio.

Nominal Scales. Nominal scales are those in which the numbers are merely substitute names for the objects being scaled. In the usual quantitative sense the number has no meaning. That is, the number does not indicate that the object has more or less of a characteristic than does any other object. Probably the most common use of a nominal scale are the numbers assigned to various members of an athletic team. Another commonly used system of nominal numbers is the Dewey Decimal System employed by many libraries.

List one or two other such systems:

1.

2.

Note: Compare your answers with those in the answer box on the next page.

Answers: Nominal Scales

1.	Sex (o = boys, 1 - girls)
2.	College Departments (1 = Anthropology, 2 = Biology, 3 = Chemistry, etc.)
3.	Makes of cars (1 = Chevrolets, 2 = Ford, 3 = Dodge, etc.)

Ordinal Scales. The lowest level of scales in which numbers have quantitative meaning are ordinal scales. As in nominal scales these numbers also differentiate between the objects being scaled; however, in addition the numbers indicate which of two objects have more of the characteristic. In some states, for example, schools are classified according to size of enrollment. The numbers assigned to the classes indicate a difference in size, e.g. Class I schools are longer than Class II schools, etc. If I were to order the institute members according to their height, assigning the number 1 to the tallest, 2 to the second tallest, etc., the set of numbers would be an ordinal scale.

What are some ordinal scales that you have experienced?

1.

2.

3.

Note: Compare your answers with those in the answer box on the next page.

Answers: Ordinal Scales

1. Order preference for a list of foods 2. List of occupations ordered by amount of income. 3. Course grades (usually)

Interval Scales. Interval scales possess all of the characteristics of nominal and ordinal scales, i.e., they use numbers to name objects and the numbers have quantitative meaning, e.g. a value of 10 may mean that the object has more of the characteristic than does the object labled 9, or 8, or 3. One important additional characteristic is added to the interval scale and that is that consecutive numbers are equally spaced along the scale, i.e. the intervals between the numbers are equal. This means that the distance between 9 and 13, for example, is equal to the distance between 7 or 3. The Fahrenheit thermometer is an interval scale. Is the centigrade thermometer? On an interval scale is 8 twice as large as 4? Why?

List two examples of interval sclaes.

1.

2.

Note: Compare your answers with those in the answer box on the next page.

Answers: Interval Scales

1. Serial numbers assigned to equidistant telephone poles.
2. Heights of airplanes above a given city.

<u>Ratio Scales</u>. The highest order of scales is the ratio scale which adds a "zero point" to the interval scale. This addition is important since it permits us to say that one score is many times greater than another, e.g. 16 is twice as large as 8 or 4 times greater than 4 or 1/3 as large as 48. Measures of distance or weight are ratio scales.

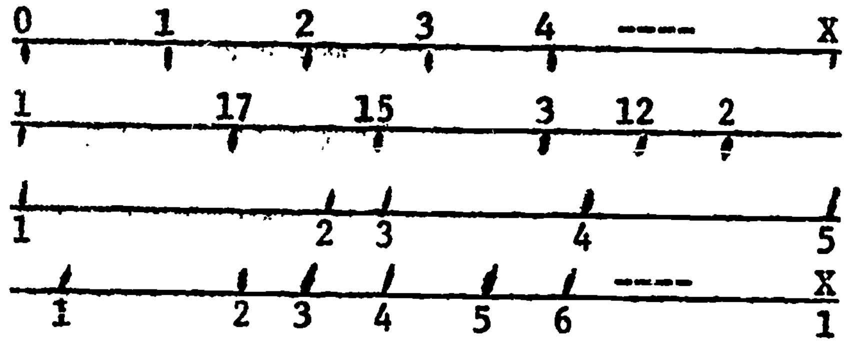
The following depict the four types of scales. Label each.

1. ______________

2. ______________

3. ______________

4. ______________



Answers: Types of scales

1. Ratio
2. Nominal
3. Ordinal
4. Interval

Most measures of psychological attributes fail to meet the criteria of ratio scales. In fact they rarely are sophisticated enough to be interval scales. They often are more than ordinal. Therefore, the data level will hereafter be referred to as being:

Class C: Nominal only

Class B: Ordinal only

Class A: Better than ordinal

Number of Samples

Comparative studies may be made using two or more samples. The distinction must be made between situations in which just two samples are being compared and in which k samples (any number of samples) are being compared.

Let's consider two sample studies first. Anytime subjects of one type are compared with subjects of another type, a two sample comparison is involved. This implies to the investigator that he must choose one group to represent one type of subjects and another group to represent the second type of subjects. Comparisons might also be made of subjects training under our set of conditions with these who trained under another set of conditions.

Graphically, two sample studies are portrayed as follows:

1.	Men	vs.	Women
2.	Sighted	vs.	Blind

etc.

When more then two samples are involved several conditions must be considered. Suppose in a given company we wish to compare attitudes towards militant Civil Rights activities held by (1) management personnel, (2) supervisory personnel, and (3) nonsupervisory personnel. Obviously we would be required to select three samples (one for each classification of personnel). A similar situation might be the comparison of effectiveness of three teaching techniques.

In each case the categorizations was made along a single factor or dimension, e.g. type of worker, and teaching method. Graphically such studies would be portrayed as follows:

1. Categories in one dimension

Teaching Method

a.

A	B	C

School

b.

Engineering	Education	Arts & Sciences

Suppose in the second example, we suspected different learning capabilities for boys and girls. We may then add a second dimension, i.e. a dimension for sex which has two categories. Addition of this dimension increases the number of samples required. In this example we have three categories of method and two categories of sex, and must, therefore have six samples (3X2).

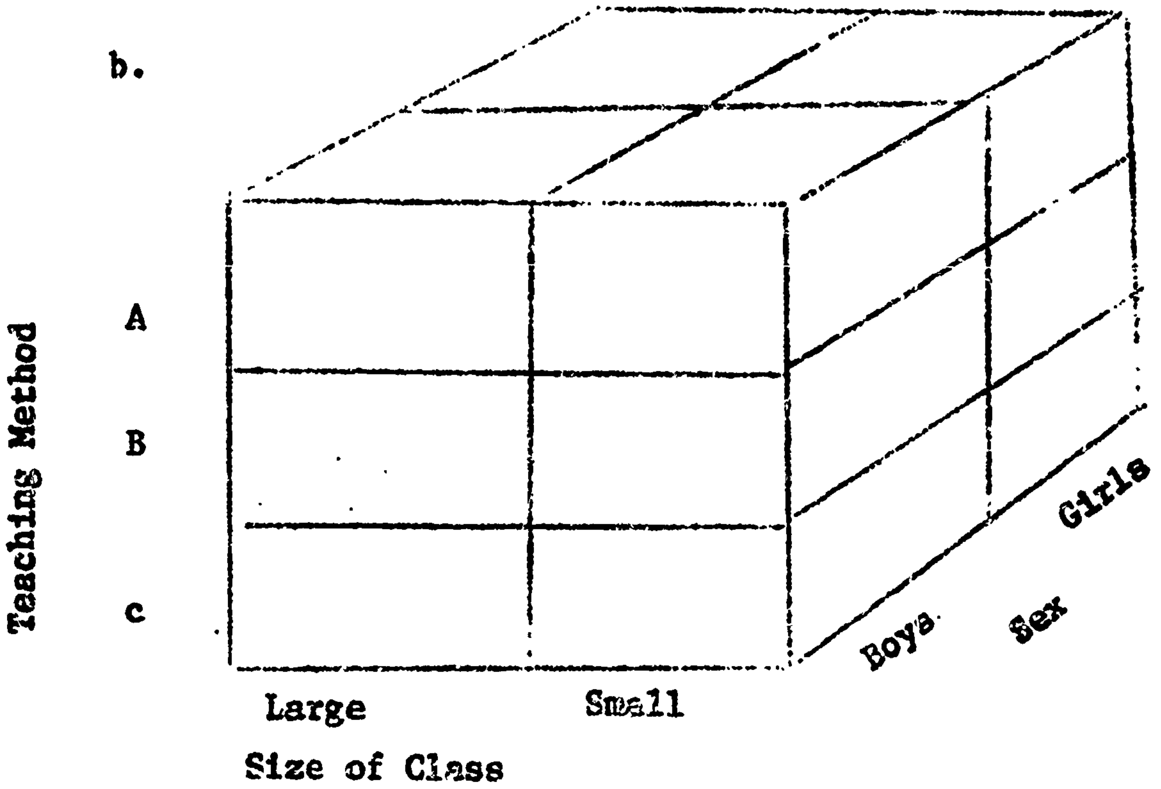
There is theoretically, no limit to the number of dimensions which might be included in a study. The point to be made here is that the number of samples increases by a multiple of [illegible] number of categories in each added dimension.

Graphically, studies in two or more dimensions are portrayed as follows:

a. Teaching Method

		A	B	C
Sex	Boys			
	Girls			

b.



Independence of Samples

Samples fall into one of two categories--they are either related or are independent. Independent samples are those which were randomly or pseudo-randomly selected. This means that any case within a population of cases had an equal opportunity to be selected. For most studies utilizing parametric statistical procedures, it is assumed that the samples were randomly selected from infinitely large, normally distributed populations.

Technically speaking, it is difficult to justify the randomness of samples used in many educational experiments. Psychologists, for example, often will assign various sections of a course they are teaching to different treatment groups and conduct their study as though each section was a random sample drawn from some hypothetical infinitely large populations. I am certain that this practice stimulates all sorts of anxiety, wrath, and scorn in the statistical purists. Such practice, however contaminated, seldom results in decisions different from those which might result from using completely random samples. In my opinion, the three assumptions (Randomness, infinite populations, normally distributed) can be horrendously violated without affecting the decision to be made.

Related samples are those which do not, for various reasons, meet criteria of randomness. An example of related samples is the comparison of pre- and post-training scores for a group of subjects. Obviously post training performance can be expected to be related to initial performance. Similarly if all subjects in a study are subjected to all treatments, the samples are related. Analysis techniques differ for studies using related and independent samples.

A Note on Hypotheses

For studies making comparisons our analysis techniques permit only for testing of hypotheses of equality, e.g., the differences between treatment groups is equal to zero (the null), the correlation between two sets of scores is-1.00. We then provide evidence which permits or fails to permit us to reject such a hypothesis.

I would like to focus these comments briefly on three formal null hypotheses.

Very frequently, research involves the comparison of sample means and variances. Far more often than not, the researcher conceives the scores of the two (or more) samples as possible

members of a single population of scores. The scores may come from persons, objects, or events that could themselves hardly be considered as a random sample from a single population. There could be a sample of boys' scores and a sample of girls' scores. The boys could hardly be considered a random sample from a mixed population, nor could the girls. But, if the scores are unrelated to the sex characteristic, (and with the null hypothesis, that would be hypothesized), then a single population of scores is a tenable hypothesis. The full null hypothesis for this general case is:

> Null Hypothesis I: The differences between the means and variances of the samples are no greater than differences due to the vagaries of random sampling from a single, normally distributed, infinite population.

Two or more samples can be analyzed at the same time with this hypothesis.

If a test of two sample variances is to be made, with no concern for the means, the null hypothesis is:

> Null Hypothesis II: The difference between the variances of the samples is no greater than the differences to be expected from the vagaries of random sampling from a single, normally distributed, infinite population.

Suppose that the second null hypothesis is tested and rejected. The only proper conclusion is that the samples do not represent a single population. Sometimes texts seem to imply that there could be a single normally distributed population that yields random samples with heterogeneous variances. This author believes that they should not so imply. To reject the second null hypothesis is to reject the applicability of the first, and the first should not be tested after the second is rejected.

If, on the other hand, the second hypothesis is tested and not rejected, that is to say, the variances are not different, then the first hypothesis can be tested. If the first hypothesis is then rejected, the differences between samples can be attributed mostly to the difference between sample means. The difference between variances still contributes a little to the difference between samples, but it was demonstrated with the second hypothesis that the difference between variances was not sufficiently great to be called statistically significant. Rigorously speaking, it is not proper to speak of a difference in means based upon a test of the first hypothesis. In practice, however, such a difference

is claimed if it has been demonstrated that the difference between sample variances is not statistically significant.

If the variances of samples are found to be different by testing the second hypothesis, there is a third hypothesis that can be used to compare the sample means for two samples. This hypothesis is based on the assumption that there are <u>two</u> parent populations with equal means. It makes no difference whether the variances of these populations (or their samples) are the same or different. This hypothesis is:

> Null Hypothesis III: The difference between means of the two samples is no greater than the differences to be expected from the vagaries of random sampling from two normally distributed, infinite populations with equal means.

It seems to this author that this third hypothesis could be used far more than it is. When the two samples have been selected because they represent two different groups, such as boys and girls, this hypothesis seems to be most appropriate. It has already been pointed out that the scores from these different groups can be thought of as representing a single population of scores, even though neither the boys nor girls can represent a mixed population of children. (This type of study in which two groups that differ in some specified way are compared as to another attribute is known as a causal-comparative study.) It seems, then, that either Hypothesis I or III can be properly used to compare means in a causal-comparative study.

<u>Causal-comparative and Experimental Studies</u>. One thing about looking at this problem with the author's bias is the nicety with which the statistical methods for causal-comparative and experimental studies are distinguished. In the experimental study, two groups, essentially identical at the outset, are compared as to some attribute after different experimental treatments have been administered. At the outset of the experiment, it is evident that the research worker feels that the samples represent a single population. He wonders if, after experimental treatment, they still represent a single population, or if the treatment has caused a difference. Therefore, he tests Null Hypothesis I. Thus, as the author sees it, Null Hypothesis I is appropriate for experimental studies and Hypothesis III is preferable for causal-comparative studies.

Here are two examples. From an apparently homogeneous community of white rats two groups are randomly selected. One group learns that a right turn on the maze, regardless of its orientation, leads to reward. The second group learns that a turn toward the window, regardless of the orientation of the maze, leads to reward. The learning scores of these two groups should be compared by testing Null Hypothesis I. Next, a group of hooded rats learns to turn toward the window to earn the reward. The "toward window" learning scores of the hooded rats are compared to those of the white rats by testing either Null Hypothesis I or III. If the researcher chooses to emphasize the possible, overall homogeneity of the learning scores Null Hypothesis I is better. If he wishes to emphasize the fact that white rats and hooded rats are different in some respects Null Hypothesis III is better. In either case, the decision to use Hypothesis I or III should be based on his conceptions of the populations that are represented, not on characteristics of the samples. Hypothesis I states that there is a single population and experimental studies start with a single population. Hypothesis III states that there are two populations and causal comparative studies distin uish between two or more samples on an a priori basis. What could be nicer than that?

Degrees of freedom. The confidence that can be placed in conclusions is related to the "degrees of freedom" associated with the analysis. The concept "degrees of freedom" is quite complex and will not be covered in any depth here. Suffice it to say that usually the degrees of freedom is one less than the number of observations, i.e. if you have 20 cases in your sample (n=20) the degrees of freedom is 19 (d.f.=n-1 or in this case d.f.=20-1=19). Maybe an example will help to clarify this.

Suppose we have a large bowl containing "peas", each of which has a number between 1 and 10 printed on it. Lets say I draw two peas (n=2) such that their sum equals 10. with the limiting factor, their sum must equal 10, I have only one opportunity to select a pea randomly. That is to say I have lost one degree of freedom or one opportunity for random selection. For example if I randomly select a pea with a 3 on it, I cannot randomly dip in to get my second pea for it must equal 7, i.e., (10-3).

It is sometimes pointed out that Hypothesis I is evaluated with more degrees of freedom than Hypothesis III, and therefore is preferable. That is getting the cart before the horse. It is true that more degrees of freedom are used with Hypothesis I. When it is assumed that each case comes from a single population every

case can be used to estimate the population paramenters. When it is assumed that each sample comes from a separate population only the cases within a sample can be used to estimate its population parameters. Since degrees of freedom are closely related to the number of cases which can be used for estimating parameters, Hypothesis I does indeed have a greater number of degrees of freedom. This is not justification, however, for using Hypothesis I, when it is not reasonable to assume that the samples represent a single population.

Unstated assumptions. As these three formal hypotheses have been stated there are no assumptions left unspecified. If the null hypotheses are abbreviated to "There is no difference between samples,: as sometimes is the case, the reader cannot be certain what the proper statistical test should be. If the reader must guess, it is best to guess that the research worker was assuming that random samples have been drawn from a single, normally distributed, infinite population. The four assumptions of the previous statement (random selection, single population, normally distributed population, infinite population) are the most common assumptions for testing hypotheses.

VIII - ANALYSIS II

Evaluation Sheet

We need to know how well the ideas and issues in this manual are communicated to you. You are the test audience for this material. To remove or strengthen the weak spots, to retain or improve the strong ones (if any), an account of your learning experience as you read these sections is crucial. Use this sheet as you study and jot down your reactions. Don't be concerned with typographical errors. We're concerned with how the message is coming through.

SECTION #________ TITLE______________________________

Give topic, paragraph or sentence, page	Importance Rate 1 2 3 4 5 Low - High	Clarity Rate 1 2 3 4 5 Low - High	Suggested Improvements

Over-all Rating of the Section: Use a 5-point Scale

Important					Understandable				
1	2	3	4	5	1	2	3	4	5
Very Low				Very High	Very Low				Very High

(Circle one) (Circle one)

PLEASE PUT ANY ADDITIONAL COMMENTS ON BACK

SECTION VIII
Data Analysis II

Table of Contents

SECTION VIII

Data Analysis II

James H. Beaird

In the previous Data Analysis section we considered questions which must be asked in order to select appropriate analysis tools and defined the parameters necessary to answer the questions. This section will "program" you through the questions and identify for you the "tools" required. Each tool is then described.

The "Program"

Selection of appropriate data analysis tools is a matter of asking (and answering) a set of relevant questions. This section is intended to permit you to identify the appropriate analysis technique for various types of research problems.

Begin by asking and answering the question, "Do I want to compare, relate, or describe?"

<u>Compare</u> - turn to A1 on page VIII-1
<u>Relate</u> - turn to B1 on page VIII-1
<u>Describe</u> - turn to C1 on page VIII-1

(A1) You said you want to compare. Now ask the question, "How many groups (samples)?"

Two samples - turn to A2 on page VIII-2.
κ samples (more than two) - turn to B2 on page VIII-2.

(B1) You said you want to relate. Now ask the question, "How many variables?"

Two variables - turn to A9 on page VIII-
κ variables - turn to B9 on page VIII-

(C1) You said you want to describe. Now ask the question, "Within which measurement level do my data fall?"

Class A --turn to A12 on page VIII-
Class B - turn to B12 on page VIII-
Class C - turn to C12 on page VIII-

(A2) You said there were two samples. "Are the samples related or independent?"

Related - turn to A3 on page VIII-2.
Independent - turn to B3 on page VIII-2.

(B2) You said there were κ samples. "What is the level of measurement?"

Class A - turn to A6 on page VIII-3.
Class B - turn to B6 on page VIII-3.
Class C - turn to C6 on page VIII-3.

(A3) You said related. "What is the level of measurement?"

Class A - turn to A4 on page VIII-2.
Class B - turn to B4 on page VIII-2.
Class C - turn to C4 on page VIII-2.

(B3) You said independent. "What is the level of measurement?"

Class A - turn to A5 on page VIII-3.
Class B - turn to B5 on page VIII-3.
Class C - turn to C5 on page VIII-3.

(A4) You said you are comparing two related samples with Class A data! Use the t-test for correlated samples.

(B4) You said you are comparing two related samples with Class B data! Use the Wilcoxon matched-pairs signed-ranks test (see Siegel, 1956, pp. 75-83).

(C4) You said you are comparing two related samples with Class C data! Use the McNemar test for significance of changes (see Siegel, 1956, pp. 63-67).

(A5) You said you are comparing two independent samples with Class A data! Use the t-test for separate group or pooled variance (recall discussion of null hypotheses).

(B5) You said you are comparing two independent samples with Class B data! Use the Mann-Whitney U-test (see Siegel, 1956, pp. 116-127).

(C5) You said you are comparing two independent samples with Class C data! Use the Chi-square test.

(A6) You said you had κ samples with Class A data! Use analysis of variance (ANOVA).

(B6) You said Class B data. "Are your samples related or independent?"

Related - turn to A7 on page VIII-3.
Independent - turn to B7 on page VIII-3.

(C6) You said Class C data. "Are your samples related or independent?"

Related - turn to A8 on page VIII-4.
Independent - turn to B8 on page VIII-4.

(A7) You said you are comparing κ related samples using Class B data. Use the Friedman two-way analysis of variance (see (see Siegel, 1956, pp. 166-172).

(B7) You said you are comparing κ independent samples using Class B data. Use the Kruskal-Wallis one-way analysis of variance (see Siegel, 1956, pp. 184-193).

(A8) You said you are comparing κ related samples using Class C data! Use the Cochran Q-test (see Siegel, 1956, pp. 161-166).

(B8) You said you are comparing κ independent samples using Class C data! Use the Chi-square test.

(A9) You said you want to relate two variables. "What level of measurement are the variables?"

Class A - turn to A10 on page VIII-4.
Class B - turn to B10 on page VIII-4.
Class C - turn to C10 on page VIII-4.
They are mixed - turn to D10 on page VIII-4.

(B9) You said you want to relate κ variables. "What level of measurement is the dependent variable?"

Class A - turn to A11 on page VIII-5.
Class B - turn to B11 on page VIII-5.
Class C - turn to C11 on page VIII-5.

(A10) You said you want to relate two Class A variables! Use the Pearson product-moment coefficient of correlation (r).

(B10) You said you want to relate two Class B variables! Use the Spearman rank-order coefficient of correlation (rho) or the Kendall rank correlation (Tau).

(C10) You said you want to relate two Class C variables! Use the contingency coefficient (C).

(D10) You said your variables are mixed! If one variable is Class A and the other Class C use biserial correlation. If one is Class A and the other Class B, convert the Class A data to Class B and use the Spearman rho or Kendall Tau.

A11 You said you want to relate κ variables where the dependent variable is Class A! Use multiple regression analysis.

B11 You said you want to relate κ variables where the dependent variable is Class B! Use the Kendall partial rank-correlation (see Siegel, 1956, pp. 223-229).

C11 You said you want to relate κ variables where the dependent variable is Class C! Use discriminant analysis.

A12 You want to describe Class A data! Use the mean ($\bar{X}$), median, or mode and variance (σ^2).

B12 You want to describe Class B data! Use the median.

C12 You want to describe Class C data! Use the mode.

The "Tools"

Analysis of Variance. The most powerful single tool for experimental studies where there is a single dependent variable is the analysis of variance model (ANOVA). The model is applicable for all data that would fall within the Class A level of measurement. It may be used anytime that assumptions of independent samples drawn from normally distributed populations are met. When comparing two independent samples ANOVA is equivalent to the t-test.

t-Tests. When data fall within the Class A level, differences between two samples may be tested using a t-test. The most appropriate t-test for most situations is the t-test for separate group variance. This is especially true for causal-comparative studies. For experimental studies, pooled variance t-tests are usually more appropriate although the distraction becomes sticky and requires close attention to some specific details. Let's consider a couple of examples.

Let's say that we have devised a new technique for teaching reading that will lead to greater retention of facts. We have a "good" test for measuring retention that yields data in Class A. To determine whether or not our new reading instruction is

superior to what we have been doing, we randomly assign our class to two groups, each of which will receive instruction under one of the techniques. The appropriate t-test is for pooled variance. The reason is that we have randomly selected two groups (samples) from a single population and have then added an experimental variable. We are trying to determine if the two samples could be expected to have been drawn from a single population.

In another study let's say that we wish to determine whether sex is a possible determinent of quantitative skills in children. Randomly selecting a sample of boys and a sample of girls from a population of tenth-graders (age 15) we again use a t-test, but this time the t-test for separate group variance is appropriate. Reasonably, we are sampling from two separate populations and are assuming that the population means are equal.

In another study we wish to learn whether any change in creativity results from training in problem solving. We design our study by measuring pretraining levels of creative ability in our sample, giving the group training, and then post-testing their behavior on the same or a demonstrated equivalent measure. Here it is obvious that some relationship between pre- and post-training performance would be expected. The t-test for correlated samples is appropriate.

Chi-Square. When data fall within Class C, the Chi-square is appropriate for testing differences between two or more independent samples. Essentially the Chi-square test is a test of independence as well as difference.

An investigator wished to determine whether college graduates were more likely to support school levies than were noncollege graduates. His findings are recorded in the following contingency table:

	Grade	Non Grade	
For	75	148	223
Against	32	111	143
	107	259	366

The obtained Chi-square value of 3.70 was not sufficiently large to reject the hypothesis that voting behavior was independent of graduate status.

In another study the investigator devised five alternative uses for a set of materials, designed to teach a given behavior. After training he noted the following distribution of those who did and did not exhibit the behavior.

	Method					
	A	B	C	D	E	
Did	16	8	8	19	21	72
Did Not	7	6	13	7	10	43
	23	14	21	26	31	115

Again the obtained Chi-square value of 7.58 was insufficiently large to suggest that acquisition of the behavior was better for any given method.

The investigator then wished to know how the trainee's attitudes toward the methods differed. He asked them whether they were positive, negative or neutral in their attitudes toward the method and found the following distribution:

	A	B	C	D	E	Total
Positive	8	5	10	13	18	54
Negative	5	6	5	6	8	30
Neutral	10	3	6	7	5	31
Total	23	14	21	26	31	115

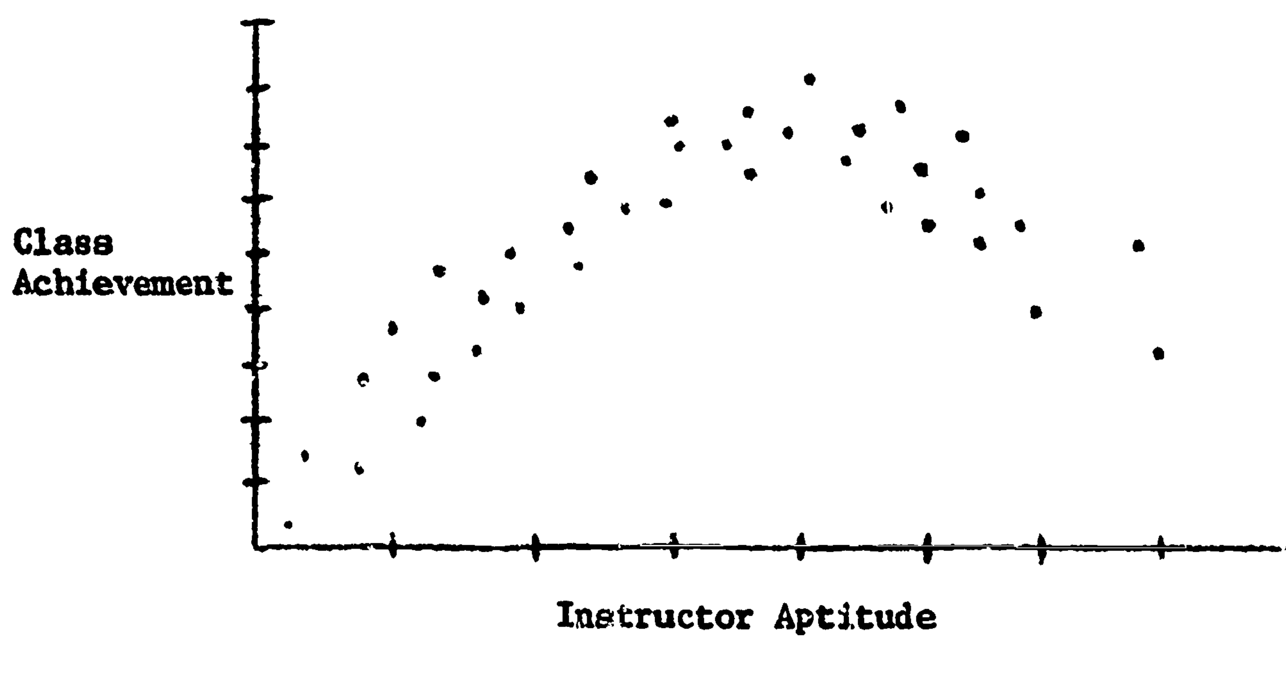
The obtained Chi-square value was 7.50, again showing attitudes independent of method.

The limiting factor of Chi-square is that only two dimensions may be categorized. Indeed this limitation holds for all nonparametric techniques.

Studies of Relationship

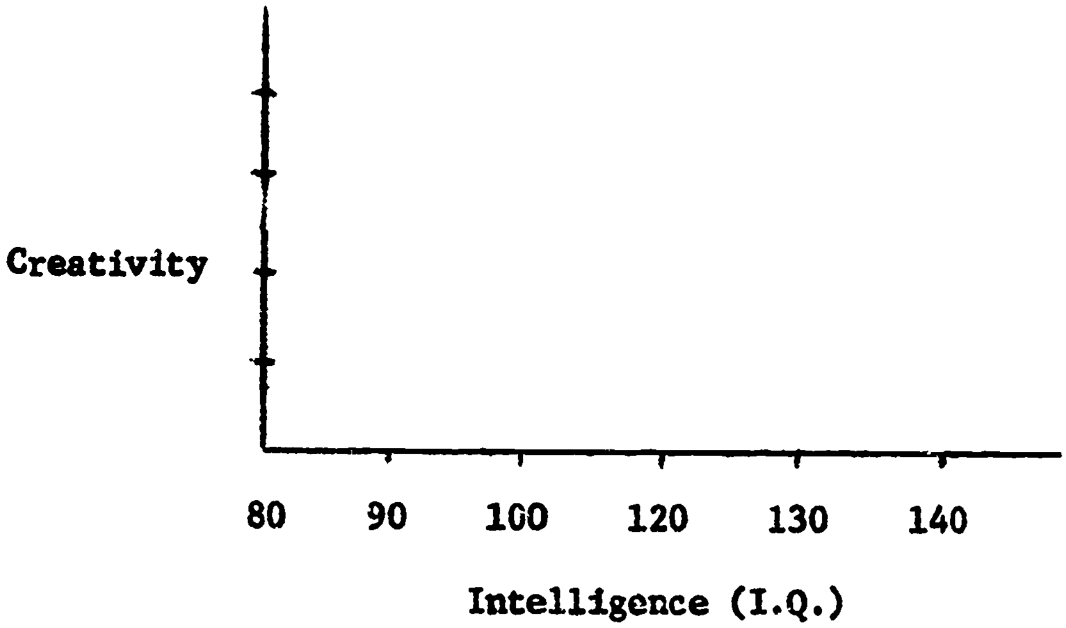
Single Order Correlation. When an investigator wishes to determine the extent of relationship between two characteristics, measures of both of which fall within the Class A level, the Pearson product-measurement coefficient of correlation (r) is the appropriate tool in most cases. The most crucial limiting factor, which might alter the use of this approach is the colinearity of the two variables. If the relationship of variables has been studied previously and evidence of linearity is present, use of r is generally safe. Without such evidence the careful investigator should first produce a scatterplot of his data, studying the plot to determine any obvious deviations from linearity. If the scatterplot provides rather clear evidence of curvilinearity, alternative correlational techniques should be used (Wert, et al., 1956).

An investigator wished to determine the relationship between quantitative aptitudes of mathematics instructors and achievement of their students. His first analysis of this relationship indicated little relationship (r = .16). The scatterplot for his data is shown below:



Would he be justified in concluding that little relationship was present?

Another investigator, in reviewing creativity research, found the statement that a rather clear relationship between creative behavior existed for those with I.Q.'s of 120 or lower but beyond that point little relationship was present. Draw a scatterplot that might illustrate such a finding.



Biserial Correlation. Occasionally an investigator wishes to determine the relationship between two characteristics, one of which is Class A and the other a dichotomous Class C characteristic. The appropriate descriptor of relationship is the biserial r. An example might be the relationship between intelligence (Class A) of unemployed adults and their political party preference (Democratic or Republican) (Class C). This technique may be expanded along the Class C variable to include 3, 4, 5, or more categories.

Spearman Rank-Order Correlation. When data variables yield Class B data, the Spearman rho is the equivalent measure to r. It requires that data variables be ordered and is able to accept ordered ties. Excessive numbers of ties tend to attenuate rho and should therefore be adjusted. An alternative method is the Kendall Tau which is used less frequently but is acceptable and to some researchers is desirable.

Contingency Coefficient. When both variables are dichotomous Class C variables, an extension of the Chi-square techniques yields a measure of relationship C. This, for example, could be applied to our previous Chi-square example of the difference in voting behavior of college graduates and nongraduates. In that analysis we found we could not reject a hypothesis of independence of voting behavior and graduate. We could further compute a C to indicate the degree of relationship. We find C = .14.

Multiple Variable Relationships. Many studies wish to determine the relationship between one variable and several other variables. The general form for such studies is the Multiple Regression Analysis model. It requires measures of all subjects on all variables and demands that the single dependent variable and most independent variables be of the Class A level. The regression model fits several types of studies. It yields a multiple coefficient of correlation, R; a prediction equation, partial coefficients of correlation (holding one or more variable constant), and finally permits analysis of relative contribution of each independent variable to the prediction of the dependent variable. The investigator is cautioned against using large numbers of independent variables in a study with this model. When the number of independent variables (predictors) equals or exceeds the number of subjects, the solution is overdetermined and R always is equal to unity. An accepted rule of thumb is to limit the number of predictions so that $K \leq N/2$, where N is the number of subjects.

Discriminant Analysis. An extension of the multiple regression model yields a model for multiple prediction of a dichotomous Class C variable. The same cautions exist for discriminant analysis as do for multiple regression. The same types of studies may be completed - the onl[illegible]ference is that the dependent variable for discrimina[illegible]ysis is dichotomous rather than continuous.

References

Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

Wert, J. E., Neidt, C. O. and Ahmann, J. S. Statistical Methods in Educational and Psychological Research. New York: Appleton-Century-Crafts, 1954.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

IX - PROPOSAL

Evaluation Sheet

We need to know how well the ideas and issues in this manual are communicated to you. You are the test audience for this material. To remove or strengthen the weak spots, to retain or improve the strong ones (if any), an account of your learning experience as you read these sections is crucial. Use this sheet as you study and jot down your reactions. Don't be concerned with typographical errors. We're concerned with how the message is coming through.

SECTION #________ TITLE______________________________

Give topic, paragraph or sentence, page	Importance Rate 1 2 3 4 5 Low - High	Clarity Rate 1 2 3 4 5 Low - High	Suggested Improvements

Over-all Rating of the Section: Use a 5-point Scale

Important				
1	2	3	4	5
Very Low				Very High

(Circle one)

Understandable				
1	2	3	4	5
Very Low				Very High

(Circle one)

PLEASE PUT ANY ADDITIONAL COMMENTS ON BACK

SECTION IX
Proposal Writing

Table of Contents

SECTION IX

Proposal Writing

Jack Crawford

The acid test of whether you have attained the objectives of this Institute is the writing of an adequate proposal. A fair degree of competence in the previous eight sections is a prerequisite. If you can't specify behavioral objectives, or can't select an appropriate design and tools for analysis, then resign yourself that any proposal you conceive will be an abortion.

But, a healthy proposal needs more than mastery of the previous material. The "more" consists primarily of certain communicative skills. The architect may have created a great design; the blueprint specifications may be accurate, clear, and detailed; the contractors capable and realistic in their estimates; etc., but somebody has to pay for the building costs. The analogous task is to convince the bank that this is a needed and worthwhile investment.

Few researchers can, out of pocket, fund the kind of research they want to do. The decision to fund or not will be made by others; and, in most instances, the decision will be based solely on the proposal. However, viewing the proposal as a mercenary medium should not obscure its other functions. The proposal can also serve:

1. to clarify ideas and detect limitations in your own thinking,

2. to communicate to peers and thus instigate useful comments and jolly criticisms,

3. as a first-draft report of the completed study. A sound proposal requires little more than a change of tense, a filling-in of data actually collected, and amplified discussion to become a complete report.

Most of us think proposals are good training devices for graduate students but not for ourselves. They are hard work.

A proposal, being a detailed plan for a research study, is addressed to three general questions:

1. Why does this study need to be done?

2. How are you going to conduct the study?

3. What will be the effects of the study when completed?

Proposal formats represent an analytic outline of the issues involved in answering these three functions. Required formats vary from agency to agency; some demanding an exhausting degree of specificity, others relatively informal. In the following discussion, the outline contains those components generally required in any proposal.

I have assumed that you have read and understood (to a moderate degree) the previous eight sections of this manual. If you are not familiar with the usual terminology in measurement, design, and statistics, you may need to refer back to the appropriate section. As a self check, you should be able to give at least a verbal definition for the following concepts:

randomization	null hypothesis
control	operational definition
stratified sample	an experimental study

Failure indicates a need for help. (Help for this instructional system.)

Objectives of This Unit

1. To identify the major components of a research proposal, their function and criteria

Desired behavior: Recall an ordered set of proposal components with criteria for each.

2. To detect and remedy certain typical weaknesses in proposals.

Desired behavior: Identify weaknesses in examples and indicate change required to approach criteria

3. To construct a sound proposal outline in a problem area of your interest.

Desired behavior: Write a good proposal, get it funded, and let us know.

The Target Audience

Most proposals are sent by the funding agency to a board of reviewers. Each of the reviewers independently analyzes the proposal and makes a recommendation. Usually, he reports reasons for his decision, and within the "recommended for funding" category may rate the proposal for priority or desirability. Decisions of reviewers are followed by the funding agency.

Since the reviewer plays such a crucial role, what are some relevant characteristics of the typical reviewer?

1. He is not a specialist in your particular problem area. He will depend upon you to acquaint him in a succinct fashion with the importance of the study and the issues involved.

2. He is extremely busy. He needs to find the meat of your proposal quickly. Obscure and ponderous writing are high on his blacklist.

3. He knows research methodology. Weaknesses and omissions will be readily detected, e.g., you can't fool him by skirting or ignoring a control problem.

4. He has an extensive knowledge of other proposals (good and bad) and of completed research. There is competition for most research funds, and the quality of the competition is increasing. Your proposal will have to rank high among those in his in-basket.

Examine the outline in Figure 1. It illustrates the sequence of topics, and over-all function of each, needed for a typical research proposal. As you read, think of research ideas of your own and relate them to the components of this outline.

The top components of the proposal will be discussed in the order of the outline. Each discussion will amplify the function of that component for the reviewer; will present some objectives

Topic	Function in Terms of General Questions Answered
1. Introduction	
Statement of the problem	What needs to be done and why?
Review of literature	What has been done that is relevant?
Objectives	What are the specific testable goals?
2. Procedures	
Design	What is the structural plan? What control will it afford?
Sampling	What population will be sampled? What size sample and how drawn?
Measurement	What will be measured? How?
Analysis	How will data from the measures be examined?
3. Product and Use	What will be the end-product of the study? What contribution could it make? How?
4. Personnel and Facilities	Who will do the study? What is their relevant competence?
5. Budget	What will each part of the study cost?

Figure 1. The Major Components of the Research Proposal

or criteria that the component should attain; and list some common weaknesses. An itemized count of proposal weaknesses, based upon a study of many proposals, is contained in Appendix C.

The Introduction

1. Statement of the problem. A terse, clear statement of the problem is a good opener. The object of this section is to present the nature and importance of the problem to the reviewer. The section consists of two parts: the major part, a description of a need; and the minor part, a brief preview of how the proposed study will relate to the need. The need for the study may rest upon its probable contribution to knowledge or its action upon society - sometimes both. If this section fails to convince the reviewer that the study is worth doing, then the entire proposal fails.

a. Criteria for judging the significance of the problem include:

(1) Generalizability - either in terms of theory, or to other populations, or to a range of practical problems, or a demonstration to be emulated, etc. The fantasy of funding agencies lightly turns toward eternal monuments to world-wide educational progress flowing from each small contract. Soberly, they want to see that you affect something more than your own corner.

(2) Theoretical contribution

(a) testing, expanding, or qualifying previous theories.

(b) adding to, creating, or otherwise contributing to new theory.

(3) Empirical contribution (similar to (2) above, except relating to observation)

(4) Practical contribution - relating to a critical social need and its solution.

(5) Methodological contribution - creating, refining, or extending an instrument or technique.

(6) Innovativeness - this criterion is part of (2), (3), (4), and (5) but is periodically emphasized by many agencies.

(7) Panic points - "do it now or it will be too late." Either the opportunity to gather the data will be gone, or the problem will have grown unmanageable, or costs will skyrocket, etc.

b. The common weaknesses in the research problem are:

(1) The problem does not strike the reviewer as significant. Probably, the description of the need has failed to grab him vitally. Some problem areas are of great significance, but the aspects stressed in the proposal are among the trivial variables involved. How can you determine the most relevant variables of a problem? One technique is to list them by such classes as: learner characteristics, teacher characteristics, instructional methods, etc. A comparison may shed light. Previous knowledge of both theory and practice are yardsticks, and the criteria mentioned above.

Clarity of expression is crucial here. Try out your presentation on colleagues. And not just old pals.

(2) A second error is presenting the problem is inflation. A small grant contract will not erase the woes of education nor will all the funds in the agency. Reviewers know this well. You can demonstrate your awareness by limiting the problem to treatable or testable scope. If this appears to involve a conflict between generality of the study and manageable limits, cut it down, then think about implications.

The slant of the proposed study may be indicated in a few sentences. Inclusion of an overview of the research direction and definitions of any esoteric terms or abbreviations to be used subsequently will warm the heart of the reviewer.

The following two paragraphs are from the problem section of a research proposal.

> Personality evaluation has been a problem of concern to psychologists for many years. The science of psychology has grown through our knowledge of the ways in which man reacts to his environment. This growth, along with man's quest to explain natural phenomena has led some men to make attempts to find some way to describe human reactions and to make predictions, even if tentative in scope, of how certain persons will react when faced with specific situations.
>
> Knowledge of human behavior has been gained through the many investigations made throughout the history of the science. This knowledge has taken numerous forms, primarily as a result of the divergent "frames of reference" used by the investigators in their attempts to explain human behavior.

List any criteria met or weaknesses in the above statement of a problem.

2. Related literature. This section should establish the base from which your study moves. Summarize the pertinent research or practices, evaluate them, and demonstrate how your project relates to them - particularly, how it contributes to progress in its specific field. This is the logical first step in proposal preparation. Often, a thorough search will uncover a new perspective to the problem or a refinement of your original approach. Typically, the search means digging through professional journals in the library. Some hints about library and other sources are given in Appendix B.

Functions of this section include:

a. Provide the non-specialist reviewer with a succinct overview of the particular area.

b. Offer some evidence of your scholarly skills.

c. Reveal that you are aware of recent developments.

d. Delineate how your study springs beyond previous work.

A lengthy list of titles or names is of no use to the reviewer. He has encountered fluent name droppers before. Instead, focus upon a few select references, really relevant to your study, accompanied by a critical analysis of their methodological limitations. This will illustrate some of your research skills as well as your knowledge of the area.

Two kinds of assertions to be avoided as the black death are:

"No relevant work has been done in the area."

or, "So much has been done and written in this area that a summary is impossible."

Reviewers feel that the authors of such statements should never have been allowed out of graduate school - or never admitted.

List functions fulfilled and/or weaknesses in the following excerpts from Related Literature components:

(1) When compared with other areas of learning the study of concept formation is retarded. The psychological literature reveals a paucity of empirical studies of thinking. Yet thinking is a topic in which educational psychologists have a strong interest.

(2) The early work of Thorndike and Edling (1912) explored this area. Hull (1941) presented his stimulating theory of the phenomena and followed this by his later extension to motivation. Relevant research has been done by Dillinger (1950), Twelker (1949), and Welch (1952). Of course, all students are familiar with the review of the field by Hickock (1938)...

3. Objectives. If the proposal is primarily directed toward an extension of knowledge, this section will be focused upon questions or hypotheses. A more applied or practical direction may shift the focus to products. The reviewer will be concerned that the objectives are achievable.

One essential requirement is that they be testable. If no possible outcome of the study can refute the objective or hypothesis, it is not testable. The research procedures must bear on the truth or falsity of stated hypotheses. This is the main reason behind our reiterated emphasis upon operational definitions and behavioral objectives. They constrain concepts and goals to face the evidence.

Many important research topics involve covert processes or categories that are not directly observable, e.g., love, intelligence, thinking. It is certainly legitimate to construct parallel lists of objectives: a theoretical and an observable set, the latter derived from, or the implied effects of, the former. In a research proposal the observables are a must. The previous sections on Behavioral Objectives and Measurement present a more adequate treatment of these issues.

In areas where theory or prior observations are minimal, questions are an appropriate form. As knowledge increases, hypotheses become more appropriate. Questions remain the typical objectives of survey studies: How many? What is the relationship between? However, the more specific the question the more it reflects the preparation of the researcher.

Disregard the nonsense frequently written about stating hypotheses in a "null" or no-difference form. If you have hypotheses, state them the way you expect the results to go. The null is an alternative (or set of alternatives) which is logically indispensible in analyzing your data. It belongs in the results section, not here.

The objectives should flow from your statement of the problem. You have stated a need. Now, in what way are you contributing to filling that need? The reviewer will then look to see if you have set up procedures to attain the objectives. In the methods or procedures section there must be a step-by-step account of what will be done to approach each objective.

Common flaws in this section include:

a. Pie-in-the-sky objectives. Vague, global goals impervious to the study's outcomes.

b. Sandbag objectives, not referred to or dealt with by the proposed procedures.

c. Hidden objectives, contained somewhere in the text of the proposal. An astute and diligent reviewer can find them _if_ he takes the time.

List requirements fulfilled and/or weaknesses in the following objectives:

(1) The purpose of the investigation is to show that the unobserved and unmeasured behavior of high school students is different from their behavior when it is observed and measured.

(2) The object of the survey is to find out what proportion of Eskimo graduates of Neeknak High School are earning less than $1,000.00 annually.

(3) The broad hypothesis is that a procedure can be followed which will lead to the initial formulation, revision, and final development of a broadly conceived theory of education based upon sociological and other relevant research findings.

Procedures

This section presents a detailed explanation of what you are going to do. It, above all other sections, is cast in behavioral terms and operational definitions: a concrete description of the project. If the reviewer gets this far, he will devote more attention to your procedures than any other section for two reasons: one, he can easily discover from this down-to-earth section just what the study may accomplish; two, procedures are a spawning bed for numerous species of error.

If possible, a short introductory paragraph indicating the over-all kind of design and the sequence of steps you will take is helpful.

1. Design. If you are using a design with a conventional name, so label it and then explain how this situational plan handles the independent variables and provides for control. The term "design" in research literature refers to the structure of an experiment rather than to a non-experimental study. You are probably almost painfully aware that the previous materials are strongly slanted toward experimental studies. Evidence from an experiment is more highly regarded than evidence derived from other approaches.

However, not all funding sources are looking for, or require, experimental approaches. A discussion of the relative merits of non-experimental techniques may be found in Kerlinger (1964). If your study is not experimental, i.e. correlational, case study, survey, etc., indicate its category and give especial emphasis to control measures that you will use.

Check back to your statement of objectives. Lay out the steps to answer each question or to test each hypothesis. A summary chart with questions or hypotheses in one column and the procedures relevant to each in an adjoining column may help.

The design description must account for any variable that might influence the results of the study. Indicate how you will control it, e.g. build it into the design and measure its effect, block it off, incorporate its measures into your tools of analysis, randomize, etc. A useful discussion of the kinds and relative merits of control measures may be found in McGuigan (1960).

No design affords perfect control. Some factors will remain as possible sources of contamination. It is the degree to which these are strained out that marks the better design. Reviewers tend to be more favorably impressed if you demonstrate your awareness of the weaknesses of your study. At least itemize the loopholes and indicate why you were unable to control for them. Sometimes you can point to cost, i.e. the enormous number of requisite control groups, time and effort required, etc. The reviewer usually has broad experience with such limitations. But, he will prefer that you point out the weak spots and why they remain.

As a reviewer, comment on this proposal design:

> A psychologist proposes to test the hypothesis that early toilet training (Head Start) leads to a type of personality noted by compulsive cleanliness; conversely, late toilet training leads to sloppiness. Previous studies have shown that middle-class children receive toilet training earlier than do lower class children. Accordingly, he asserts he will select two groups, one of middle-class and one of lower-class children. He will give both groups a finger-painting task and compare the amount of smearing and how many times they wash the paint off.

If you have cited any design weaknesses, point out how to correct them.

2. Sampling and measurement. Because you wish to generalize any findings beyond the particular subjects in your study and/or beyond those particular days and places when the investigation occurred, most research studies are perceived as operating with a sample and generalizing their conclusions to a larger population.

This target population must be adequately described. The reviewer should be able to tell readily who is included and who is not included in the population. To a large extent your choice of sample determines the population to which the study applies. Logically, the sequence would be to define the population and then specify how a random sample, or some modification of a random sample, would be selected. In instructional research you are often constrained to study the students available, so that you must construct the relevant features of the population from the characteristics of your sample. List those subject characteristics which previous knowledge indicates may influence the behavior you are measuring. Omit characteristics which make no effect on the measure. Color of eyes (within the normal range of colors) rarely affects measures of learning.

In theory, in the proposal, and actually, you will probably define the population first - in terms of relevant characteristics - that is, characteristics probably having measurable effects on the dependent variables. Then, show how you will select a sample. The reviewer will be concerned whether your method of sampling tends to systematically bias the measures you take. Point out your efforts at randomization or stratification. If you have to use intact classes, schools, or other inclusive groups be sure you have clearly stated this.

Random sampling procedures are highly rated by reviewers and for most purposes yield logically defensible results at relatively low cost. Meticulous reviewers tend to look for 3 kinds of randomization procedure (in order of priority):

(1) Selecting the subjects from the population

(2) Assigning the subjects to groups, e.g. in a form group study placing them in Group I, II, III or IV

(3) The, and this is a separate step, assigning the the treatments called for in your design to the groups.

Generally, your sampling concerns relate to the people whose behavior you are studying. Occasionally you may have an interest in sampling from a variety of treatments. Get some design consultant assistance on this. It does offer power in the conclusions attained.

SAMPLE SIZE. Specify the size of each group in the study. Three factors should be considered in determining the number of subjects involved.

1. The degree of effect that interests you. If you desire or demand that your instruction or treatment makes a noticeable effect on a small group (say 20 students), use small groups. If it makes only a tiny effect, you can meet all statistical criteria but you will need larger groups.

2. The cost involved. Larger numbers, if not easily available, may exponentially inflate your budget.

3. Statistical considerations. Previous knowledge of <u>any</u> consistent effect can be translated into the sample size required to reach the desired statistical conclusion - but <u>only</u> when the effect is known to be in the appropriate direction and its extent as well as an estimate of variability has been measured. This is one by-product of pilot studies. Almost any statistics book or graduate student in a statistics course can show you the translation procedure.

If possible, plan to use equal numbers in each group. If you can't assume equality, get some consultant help (statistical) to point out how you will deal with unequal numbers. Reviewers like to see that you have worried about this.

SCHEDULE. A realistic time schedule adds weight to your proposal. Graphic or flow chart representations are easiest for the reviewer to follow. Large scale proposals are often accompanied by a PERT chart. This system, Program Evaluation Review Technique, takes a bit of study but has a high payoff in increased accuracy of your planned sequence of actions.

If you are not experienced in the kind of research you propose, allow more time than you anticipate. Copies of materials and instruments will refuse to appear on time, subjects will hide or get busy with life or death urgencies, your assistant will misread the instructions and you'll be searching for a new group of subjects as well, etc., etc. And remember, all funding agencies require a report. This takes secretarial and duplication time.

Usually the funding agency will indicate, in advance, a starting date if the project is approved. Allow for a delay even in this.

3. Analysis. This component describes how you will examine the data to obtain evidence bearing on your objectives. The appropriate examination tools are determined in part by the objectives, in part by the design, and in part by the class of data. The previous sections of this manual, VII and VIII, are useful guides. (1) Name the analysis tool, (2) show why it is appropriate, and (3) indicate how the product of the analysis will bear on your objectives.

If you have a proposal constructed to this point, and are still a bit cloudy about the appropriate analysis, seek help. In a small institution you can usually find some staff member in mathematics, psychology, economics, or even education who can and will help for the price of coffee and a sympathetic ear to his troubles with the administration. If possible, seek out one who has been teaching statistics regularly and recently. If you can't find local assistance, call the nearest Regional Office of the USOE. Addresses are in Section X. They will refer you to the nearest specialist or provide direct assistance.

This is the place to set out the null hypotheses. A useful technique is to place the research hypothesis of interest in one column, the alternative null in another, and in a third column, the kind of result from the analysis that will indicate rejection of the null to be the logical decision.

Frequently you will be unable to cite the appropriate tools because unknown, as yet, dimensions of the forthcoming data will determine what tool to use. Spell out the contingencies as you see them. Show alternative plans of analysis. Reviewers wax warm and beneficent when they perceive you have anticipated problems and have planned coping strategy.

Product and Use

The purpose of this section is to point up what contributions the completed study will make and how. The final report itself may be suitable for dissemination through other channels. A journal article may be a reasonable expectation.

Both effects on the body of knowledge and on working practitioners may be estimated. Suggest how dissemination to the latter may be effected. Often the study will generate by-products such as evaluation instruments, instructional materials, films, etc. These may be highly useful to others.

Perhaps a significant contribution to your own institution is a likely yield. Even the effects on the research team or the individual researcher in terms of professional experience and increased capabilities.

But, keep it brief.

Personnel and Facilities

Name the people who are to work on the study. Briefly indicate experience <u>relevant</u> to the research. If you haven't any, cite attendance at this institute. Evidence of interest in pursuing this kind of research is appropriate if experience is lacking. Don't deliver a eulogy to a bunch of nice guys. Reviewers loathe nice guys.

As a last ditch resource, consider acquiring the services of an experienced researcher as a consultant to the project. He will insist on being paid and this will add to the budget. Spell out clearly just how he will serve the project. Retain the decision function and allow him to advise and recommend.

If you have special facilities that will aid or are required by the project, list them. Libraries suitable for graduate work, computers, training programs, secretarial services, and office space may be relevant.

Budget

The easiest way to get a feel for proper budgets is to peruse some approved budgets for projects granted by the same agency. The agency will have sent you as part of the application form, probably a rather detailed budget schedule. Use some worksheet facsimile and construct a sample budget. Show it to an experienced researcher. Usually, he can quickly scan it and give you some useful suggestions. Your own fiscal officer may help. The institution may have to contribute costs you hadn't anticipated unless you provide for this. Show each facet of your institution's contribution from staff time, secretarial services, materials, etc.

The usual error of first efforts at budget construction is to underestimate and omit. If your procedure is detailed, the description can be used as a guide. Check every step including preparation of the final report.

Don't be overconcerned with this component. Budgets on approved projects are often changed through negotiation with the funding agency. Consider the proposal budget as a good first draft. Many first rate projects have to be re-budgeted later.

Selecting the Funding Agency

Supporting sources are so numerous that this entire manual could not list them. A guide to sources of information about funding agencies is contained in Appendices C and D of this section. Section X contains some guides from the United States Office of Education your most probable funding source.

For your first project, don't overlook your own institution. The president of a small college can often dig up a little seed money for a pilot project.

Agencies will send you their requirements, formats, and usually some indication of their interests. Often a short letter outlining the project in an informal way will help them to decide if it is worthwhile to submit a detailed proposal.

The Research Marketplace

The preference of agencies for different kinds of projects changes over time. I have stressed the orientation that if you are interested in a problem area, work up your ideas; then locate the appropriate funding source. A contrasting approach is followed by some researchers. These social science street walkers try to sound out agencies as to preferences, then write up proposals in that area - wherever the demand lies.

Some compromise with reality demands is inevitable. However, the increasing multitude of funding sources lends assurance to the idea that a good proposal will get supported (sooner or later).

What To Do When Your Proposal Is Rejected

After you have decided to change your specialization, written and torn up several suitable rejoinders, etc., find out <u>why</u> they turned it down. Most agencies will sent you a summary of reviewer comments. Evaluate them. If the critiques are apt, re-do the proposal. Often the agency will indicate if they think re-submission is suitable. Or if the proposal looks good to you, submit it to another agency. Some good projects

have gone through three rejections from different agencies, then been funded, and subsequently acclaimed as hallmarks of progress. To gain perspective, talk to any publishing author.

Good luck. A complete proposal is contained in Appendix A. This was approved, receiving high ratings from reviewers, despite some weaknesses (which will be obvious to you).

Reference List

Kerlinger, F. S. Foundations of Behavioral Research. Holt, New York, 1964.

Krathwohl, D. B. How to Prepare a Research Proposal. Syracuse University Bookstore, New York, 1965.

McGuijan, F. J. Experimental Psychology: A Methodological Approach. Prentice-Hall, New Jersey, 1960.

Menzel, D. H., et al. Writing a Technical Paper. McGraw Hill, New York, 1961.

APPLICATION TO THE COMMISSIONER OF EDUCATION, U. S. OFFICE OF EDUCATION DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE, FOR A GRANT TO SUPPORT A RESEARCH PROJECT UNDER THE PROVISIONS OF TITLE VII OF THE NATIONAL DEFENSE EDUCATION ACT OF 1958 (P. L. 85-864)

Submitted by: TEACHING RESEARCH
A Center for Research on Teaching
Oregon State System of Higher Education

Address: Monmouth, Oregon

Telephone number: SKyline 7-1421, Ext. 73

Initiated by: Dr. Bert Y. Kersh, Associate Research Professor
TEACHING RESEARCH

Transmitted by: Dr. Roy E. Lieuallen, President
Oregon College of Education

Date: July 26, 1961

Bert Y. Kersh

Roy E. Lieuallen

I. Project Title.

Directed Discovery vs. Programed Instruction: A Test of a Theoretical Position Involving Educational Technology.

II. Problem.

Increasingly more importance is being placed on the process of learning by "directed discovery". Representative of this increasing emphasis is the newer techniques and materials presently under development by the UICSM (University of Illinois Committee on School Mathematics) directed by Professor Max Beberman. By requiring the learners to discover for themselves new relationships in mathematics, it has been demonstrated that high school students are capable of learning advanced concepts more effectively than otherwise. The act of discovery itself is considered to be the primary reason for the success of the new mathematics curriculum (Beberman, 1955; Hendricks, 1961).

Paradoxically, increasingly more importance is also being placed on the process of learning by programed instruction ("teaching machines"). "Discovery" by the student is minimized in programed instruction which characteristically presents the material in such small steps that the learner does not have to search far to find an answer. This is especially true when prompts and cues are added. Yet, the research evidence to date indicates that programed instruction is at least as effective as more conventional methods of instruction and may be far more efficient in terms of teaching time.

The fact that both the "discovery" and the "programed" instructional methods are presently in the forefront of attention serves to revive an older, unresolved research problem in a new context, and with greater potential for solution. The problem is to determine which process of learning is superior, a highly directed (formal) learning which places the learner in a position of complete dependence on the teacher, or non directed (informal) learning in which the learner must rely almost completely on his own cognitive capabilities. Advocates of the non-directed, informal (hereafter called "discovery") process claim a number of advantages, most of which are included in a recent article by Bruner (1961). Bruner hypothesized that learning by discovery benefits the learner in four ways: (1) it increases the learner's ability to learn related material, (2) fosters an interest in the activity itself rather than in the rewards which may follow from the learning, (3) develops ability to approach problems in a way that will more likely lead to a solution, and (4) tends to make the material that is learned more readily accessible in memory - that is, easier to retrieve or reconstruct.

Research evidence is not entirely supporting of the claims of Beberman and Bruner. The more recent evidence suggests that learning by discovery does not necessarily benefit the learner directly in terms of retention and transfer, but it does foster interest in the task (Kersh, 1958). When interest is generated, the learner tends to continue the learning process autonomously beyond the formal learning period. As the results of his added experience, the learner then raises his level of achievement, remembers what he learns longer and transfers it more effectively. However, sufficient interest to continue learning evidently does not appear unless the learner expends intensive effort without help for an extended period (Kersh, 1961)

In other words, the evidence referred to above suggests that learning by dis-

covery is superior to highly directed, formalized learning only in terms of increasing student motivation to pursue the learning task. Learning with direction is far more efficient in that the student learns more during a given period of time and, when the direction is pertinent, comes to understand what he learns more completely. Obviously, programed instruction may be allied with highly directed, formalized teaching methods.

The explanation for the elusive drive generated by independent discovery is not evident, but several have been offered, including the Zeigarnik effect of superior memory for unfinished tasks and the Ovsiankina effect of resumption of incomplete tasks (Kersh, 1961). Neither of the latter is entirely adequate because they simply describe the conditions under which the motivation may be expected to appear. Actually, the motivating effect may appear even when the tasks are completed. Another explanation, and the one which is of primary concern in this present proposal, is that the motivating effect is learned through a process of operant conditioning. By this theory, the learner, who is forced to discover the solutions to problems without help, engages in a kind of behavior often described as "searching". The searching behavior is reinforced by the teacher who is monitoring the learning process, and by the learner's own successful progress towards a solution. This explanation would support the claims that Bruner makes regarding the increase in the learner's ability to learn in related areas, to solve similar problems, and to continue the searching behavior beyond the formal learning period whether or not there are any extrinsic rewards involved. The theory also fits well with the research evidence which suggests that the motivating power of learning by discovery does not appear unless the learner engages in such searching behavior over a relatively long period of time.

If the above explanation is true, then it should not be necessary to rely completely on the less efficient, time-consuming discovery techniques. In short, it should be possible to employ the techniques of programed learning to achieve the same desirable benefits which may accrue from learning by discovery. Many of the techniques of programed learning are based on the operant conditioning model and, through the use of appropriate educational media, are adaptable to any teaching objective which may be translated into behavioral terms.

The purpose of this present project is to conduct an experiment to test the theory that a teaching (machine) program which is specifically designed to develop and foster the kind of searching behavior that characterizes the discovery process, will produce equally as well the desirable benefits to the learner which are usually attributed to discovery techniques alone.

If the experimental evidence substantiates the theory, it will warrant further study. The development and eventual application in the classroom of efficient and effective techniques for using the new media associated with programed learning will no doubt follow.

III. Objectives

The present project has two objectives:

1. To develop a "programed" modification of an existing unit of instruction (The Associative Law) modeled after the UICSM "discovery" method, and adapt it to a group-pacing technique which provides individualized feedback to the learner (see detailed description under Procedure, below).

2. To conduct an experiment with the programed method designed to test the following hypotheses:

a. Students taught by the programed modification (Objective 1, above) achieve the learning objective more rapidly than students taught by the discovery method.

b. Students taught by the programed method spontaneously employ the learned material as frequently after the formal learning period as students who are taught by the discovery method.

c. Within a period of eight weeks following the formal learning period, students taught by the programed method are able to recall the learned material from memory and apply it in the solution of mathematical problems as well as students who are taught by the discovery method.

d. In a task of new learning following within 24 hours after the formal learning period, students taught by the programed method reveal by their written work and self reports that they employ the same or the equivalent techniques of independent discovery and problem solving as students taught by the discovery method.

IV. Procedures.

The project will be conducted in two phases: (1) materials development phase, and (2) the experiment. During the materials development phase, the necessary modifications to the discovery materials will be produced and the techniques developed for using them. During the same phase, the teachers for the experimental classes will be trained in the use of the materials and techniques. The first phase is expected to last approximately eight months. A more detailed description of the programed materials and techniques is given below.

The Experiment

Design. Two experimental groups of elementary school children (grade 5) will be taught the Associative Law in different ways: one group by the Discovery method, and the other by Programed instruction. Two different instructors will be involved. Then the experiment will be replicated with two new groups of children. This time the same two instructors will teach, but they will exchange methods of instruction.

The counter-balanced design is diagramed below:

Instructional Method

Experiment	Discovery	Programed
I.	Teacher A n = 15	Teacher B n = 15
II.	Teacher B n = 15	Teacher A n = 15

N = 60

Selection of subjects. A total of 60 fifth graders will be selected from the Monmouth Elementary School (and neighboring schools) on the basis of a pre-test of their knowledge of those number concepts and arithmetical operations that are pre-requisite to learning the Associative Law. To better insure that the children will be capable of learning the required concepts, their records of scholastic aptitude and achievement in arithmetic will be examined. Only children within prescribed limits of ability and achievement will be accepted. Thereafter, the group will be sub-divided into four groups of 15 by random processes, and each group will be assigned to one of the two experimental groups.

The actual instruction will take place on the campus of Oregon College of Education. The physical facilities will be the same for each group with the exception of the special equipment used with the Programed Group.

Experimental techniques. The Programed Group will be taught in the specially equipped laboratory classroom of Teaching Research, called the Teaching Research Laboratory. The laboratory is equipped with a 25-student-station TELETEST Communication System (see appendix). This system will permit the classroom instructor to present programed materials to the class and to provide individualized feedback to each student immediately after the student signals that he has made his response. The fact that each student has made a response and the particular response made is recorded in coded form immediately on an IBM (International Business Machine) type card at the instructor's station. The instructor, therefore, has the capability of determining, at any point and time, precisely what proportion of the class is ready to go on, and which individuals are progressing satisfactorily and which need help.

The programed method of instruction, using the apparatus described above, will not be the self-instructional, individually-paced approach which typically characterizes the teaching machine. This present technique uses the programed materials as an integral part of the more conventional approaches to classroom instruction. A teacher is present at all times, and group-pacing techniques are used. The teacher conducts the class by providing information, assigning problems, interacting with individuals, etc., as required. The unique characteristics of the proposed programed method are the following: (1) The teacher will follow a "program" of questions, problems, presentations, etc., which will employ such techniques as "vanishing" cues, small steps, branching, etc., as may be appropriate. (2) Each student will record his needs and reactions (e.g., "progressing satisfactorily", "need help", "repeat"), or his solutions to problems through the use of the communication system described above. (3) The teacher will receive and collate the students' reactions to questions and problems almost immediately and adjust his program accordingly. (4) The program of problems and tasks will require the students to employ problem solving and discovery techniques with increasing frequency and on increasingly more difficult tasks. In the beginning the program will be highly directive in "teaching" the students principles for solving mathematical problems and using the inductive method.

The Discovery Group will be taught by techniques and materials modeled after the UICSM course of study. The materials for this present study were developed for the fifth grade level at Oregon College of Education during 1960 and 1961.

The formal learning period is expected to last approximately ten hours distributed over two weeks. During the learning period, each subject will be brought to the same level of achievement as indicated by pre-established criteria. The learning criterion will be in terms of each student's performance on problems requiring the application of the mathematical law to be learned. Since both teaching methods use group-pacing techniques, all students will complete at approximately the same time. However, in the Discovery group especially, considerable freedom is allowed the individual to learn at his own rate, so every student will be given the opportunity to demonstrate his achievement on a short written examination whenever he and the instructor agree that he may be ready. When an individual demonstrates that he has achieved the required level, he will be excused from further attendance in the experimental classroom.

The entire experiment will be repeated twice with a three-month time interval as outlined in the experimental design, above.

Test of new learning. After each individual student completes the initial program of instruction, and within 24 hours thereafter, he will be taken aside individually and given a new task to learn. The task will be that of discovering a novel rule for adding a series of odd numbers. The rule is usually discovered within a period of 30 minutes. During the new learning period, voice recordings will be taken of the subject's verbal report of their thought processes, and any scratch work will be retained.

Post-test of recall. A post-test consisting of problems similar to the ones used during the learning period will be given to each subject within eight weeks following the formal learning period. One third of each group will be administered the post-test two days after the formal learning period, another third will be given the post-test after two weeks, and the final third after eight weeks. At the same time the post-test is given, each subject will be asked to fill out a questionaire on his use of the learned mat rial during th intervening period. They will be asked the number of times the rule was used spontaneously, and the purpose and occasion in each case.

Analysis of data. The data will be analyzed by comparing the two experimental groups in terms of time to learn, techniques employed during the new learning task, performance on the post-test, and use of the learned material during the period intervening between the learning period and post-test.

Comparisons in rate of learning (Hypothesis a, above) will be in terms of the number of class sessions (or hours) to complete the learning task. Chi square will be used to test the significance of the obtained differences against a theoretical position of equal time to complete.

Comparisons in average frequency of use of the new learning (Hypothesis b) will be based on the questionaire data in which each student will be asked to estimate the total number of times he used the information during the intervening period (together with corroborating information). If the frequency data meet the essential distribution and variance requirements, the mean differences between groups will be tested by the *t* test; otherwise, a non-parametric statistic such as chi square will be employed.

Hypothesis c, pertaining to memory for the task, will be tested with the post-test data. The tests will be scored on a pass-fail basis, since the concern is with memory for a mathematical law - not with computational accuracy or specific procedures which may contribute to overall test variance. Consequently, the chi square technique may also be employed in this analysis.

The data pertaining to the last hypothesis (d) will be in the form of frequency distributions of the techniques of problem solving and discovery used by the experimental subjects. The differences in the distributions for each group will be tested statistically by chi square, if suitable.

Approximate time schedule.

The developmental phase of the project will last approximately eight months. The task of programing involves several try-out and revision cycles, each of which is slow and tedious.

The experiment will require eight months to conduct and two months for analysis and reporting.

Expected end product.

The experiment will provide new evidence on the relative effectiveness of the methods of programed learning and discovery particularly as involves the theoretical position outlined above.

Publication plans.

The results will be submitted for publication in the Journal of Educational Psychology. A discussion of the methods employed together with their respective theoretical basis will be submitted for publication in the Mathematics Teacher and related periodicals.

V. Personnel

Project Director - Bert Y. Kersh, Ph.D. (Educational Psychology, University of California, 1955) present position; Associate Professor on the permanent research staff of TEACHING RESEARCH, Oregon State System of Higher Education. Research experience: Consultant in Educational Research, University of California, 1952-1954; recipient of three research grants from the Graduate School of the University of Oregon, 1955-1959; Human Factors Scientist, System Development Corporation, Santa Monica, California, 1959-1960.

Director of Curriculum and Instruction - Vern D. Hiebert, M.S.,(University of Illinois, 1959) Assistant Professor of Mathematics, Oregon College of Education. Related experience: Developed the grade 5 experimental materials and techniques at Oregon College of Education which are modeled after the UICSM course of study and employ the discovery method. Completed a graduate course in mathematics instruction at University of Illinois taught by UICSM personnel. Five years teaching experience in public schools.

Experimental teacher I. Vern D. Hiebert, M.S. (above)

Experimental teacher II. To be selected from the staff of the Monmouth Elementary School, the laboratory school for Oregon College of Education. He will be instructed in the experimental techniques by Professor Hiebert.

Research Assistant - Donald Toevs, M.A., Research Instructor, TEACHING RESEARCH, Oregon State System of Higher Education. Related experience: Research assistant under the supervision of Dr. Jack V. Edling, Director of TEACHING RESEARCH, 1960-62. Majored in Mathematics and Education.

Secretary - Mrs. Neil Amerman, Secretary to the staff of TEACHING RESEARCH, Oregon State System of Higher Education.

VI. Facilities

The resource personnel, office and equipment of Teaching Research, A Center for Research on Teaching, Oregon State System of Higher Education. An agency of the State Board of Higher Education, Teaching Research utilizes the full resources of the seven campuses in Oregon's unified State System, including modern motion picture production facilities, advanced electronic computer systems, completely equipped laboratory schools, and a distinguished faculty of behavioral scientists.

The resources of the Monmouth Elementary School, the Laboratory School on the campus of Oregon College of Education. The Teaching Research Laboratory is located in the Monmouth Elementary School.

VII. Duration

Total amount of time required: 18 months.

Beginning: January 1, 1962

Ending: June 30, 1963

BUDGET

Category	OE Funds Requested	OSSHE Funds Contributed
I. Direct Costs		
Personnel		
Project Director (.25 FTE, 18 mos. 1st 6 mo. paid by State, next 12 mos. Fed. Funds)	$ 2,700	1,300
Director of Curriculum & Instruction (.25 FTE, 6 mos.)	850	
Experimental teacher (one only, 1.0 FTE, 1 mo.)	600	
Research Assistant (.50 FTE, 15 mos.)	2,400	1,200
Secretary (.25 FTE, 12 mos.	425	425
Programing consultant (one only, 7 days @ $50/day)	350	
Supplies and Materials		
Programing materials (mimeo., 35 mm. film, etc.)	75	
Instructional materials		25
Office supplies		50
Other direct costs		
Payroll assessments (6% salaries)	419	175
Depreciation on TELETEST System (Total value of equip. $4,500 - Depr. est. for 1½ yrs. for proportion of time equip. used on project)		100
Communications		50
Travel (for experimental subjects)	50	
Total direct costs	7,869	3,325
Indirect costs (25% of direct costs)	1,967	none
	$ 9,836	

(cont'd. on next page)

BUDGET (cont'd.)

Summary

Office of Education funds requested	$9,836
Oregon State System contribution	3,325
Total cost of project	$ 13,161

Other support

This proposal has not been submitted to any other agency or organization.

The research proposed herein is not an extension of or addition to a project previously supported by the Office of Education.

Estimated cost to Federal Government by fiscal year.

Category	Federal Runds Requested
Fiscal year 1962	909
Fiscal year 1963	8,927
Total cost (all years)	$ 9,836

BIBLIOGRAPHY

Bruner, J. S. "The act of discovery", Howard Educational Review, Winter 1961, 31, 21-32.

Beberman, M. An emerging program of secondary school mathematics. Cambridge: Harvard University Press, 1958.

Hendrix, Gertrude. "Learning by discovery", The Mathematics Teacher, 1961, 54, 290-99.

Kersh, B. Y. "The adequacy of 'meaning' as an explanation for the superiority of learning by independent discovery", Journal of Educational Psychology, 1958, 49, 282-92.

Kersh, B. Y. "The motivating effect of directed discovery", (accepted for publication, Journal of Educational Psychology, 1961).

Appendix B

Shortcomings found in study-section review of 605 disapproved research grant applications, April-May 1959. All percentages are to the base number 605.

No.	Shortcoming	%
	Class I: Problem (58 per cent)	
1	The problem is of insufficient importance or is unlikely to produce any new or useful information.	33.1
2	The proposed research is based on a hypothesis that rests on insufficient evidence, is doubtful, or is unsound.	8.9
3	The problem is more complex than the investigator appears to realize.	8.1
4	The problem has only local significance, or is one of production or control, or otherwise fails to fall sufficiently clearly within the general field of health-related research.	4.8
5	The problem is scientifically premature and warrants, at most, only a pilot study.	3.1
6	The research as proposed is over-involved, with too many elements under simultaneous investigation	3.0
7	The description of the nature of the research and of its significance leaves the proposal nebulous and diffuse and without clear research aim.	2.6
	Class II: Approach (73 per cent)	
8	The proposed tests, or methods, or scientific procedures are unsuited to the stated objective	34.7
9	The description of the approach is too nebulous, diffuse, and lacking in clarity to permit adequate evaluation	28.8
10	The over-all design of the study has not been carefully thought out	14.7
11	The statistical aspects of the approach have not been given sufficient consideration.	8.1
12	The approach lacks scientific imagination.	7.4
13	Controls are either inadequately conceived or inadequately described.	6.8
14	The material the investigator proposes to use is unsuited to the objectives of the study or is difficult to obtain.	3.8
15	The number of observations is unsuitable.	2.5
16	The equipment contemplated is outmoded or otherwise unsuitable.	1.0
	Class III: Man (55 per cent)	
17	The investigator does not have adequate experience or training, or both, for this research	32.6
18	The investigator appears to be unfamiliar with recent pertinent literature or methods, or both.	13.7
19	The investigator's previously published work in this field does not inspire confidence.	12.6

20	The investigator proposes to rely too heavily on insufficiently experienced associates.	5.0
21	The investigator is spreading himself too thin; he will be more productive if he concentrates on fewer projects.	3.8
22	The investigator needs more liaison with colleagues in this field, or in collateral fields.	1.7
	Class IV: Other (16 per cent)	
23	The requirements for equipment or personnel, or both, are unrealistic.	10.1
24	It appears that other responsibilities would prevent devotion of sufficient time and attention to this research.	3.0
25	The institutional setting is unfavorable.	2.3
26	Research grants to the investigator, now in force, are adequate in scope and amount to cover the proposed research.	1.5

Source: Allen, Earnest M. (NIH) "Why Are Research Grant Applications Disapproved?" Science, Nov. 25, 1960. p. 1533.

Appendix C

Where to Obtain Information on Funding Sources and Procedures

1. Guide To Support Programs for Education.
Education Service Press, Visual Products, 3M Company,
St. Paul, Minnesota. 1967 (2nd. ed.) Price $12.00.

This is a worthwhile investment for the novice. Deals with principle federal laws aiding education, aid from business and foundation sources, and provides helpful hints generally on how and where to obtain support for research, development, or training programs.

2. College and University Reporter (Topical Law Reports). Commercial Clearinghouse, Inc., 420 Lexington Ave., New York. Price $455 yearly subscription, includes weekly bulletin on recent developments in Washington, D.C. in legislation relating to education and research.

This is an expensive reference source, but the services provided to the subscriber are well worth the investment. Two large loose-leaf books are included, containing detailed and very much up-to-date information on all major developments in the field of education. The volumes are revised each week when the company (CC) mails supplementary loose-leaf pages to all subscribers. Included, too, are weekly bulletins dealing with recent developments in Washington, D.C., and a copy of each law or pending law in both the House and Senate pertaining to education. This is a must for larger research organization.

3. Programs and Services: U.S. Dept. of Health, Education and Welfare.
Supt. of Documents, U. S. Printing Office, Washington, D.C., 20402, Price $2.00.

This is a useful and comprehensive guide to federal programs administered through the Dept. of Health Education and Welfare. In light of the coverage and relative cost, this is highly recommended as a valuable reference source.

4. Office of Education Support for Research and Related Activities.
Dept. HEW. Washington, D.C. Free (?) on request.

This 22-page manual provides summary information on patterns of support and application procedures through the U.S. Office of Education. It is a helpful reference guide for anyone involved in educational research.

5. Grant Data Quarterly (1st four issues in 1967). Academic Media, Inc. 10835 Santa Monica Blvd., Los Angeles, California 90025. Price: Single subscription $35.00 (10% off on 2 or more).

The first four issues present detailed information on government support programs, business and professional organization support programs, and foundation support programs. This quarterly would be valuable as a reference source for college libraries, or progressive departments contemplate a substantial volume of research and development activities.

Appendix D

Partial List of Fund Sources for Educational Research

A. Cooperative Research Program (HEW)

References: See College and University Reporter (C.C., Inc.), at 1651; also 8911 and 8401. See Programs and Services of the U. S. Department, HEW, pp. 223-224, and 231-232.

1. The Cooperative Research Programs[1] administered by the U. S. Office of Education include support for both basic and applied research, demonstration projects, and curriculum improvement projects. Funds were also authorized for the creation of research and development centers, educational laboratories, and for various developmental activities.

2. Small contracts. Perhaps of greatest significance to the researcher just beginning a career in educational or behavioral research is the small contracts program. This program is intended to provide support for small-scale research or development projects which require less than $10,000 in federal assistance.

   The program "supports experimental research, surveys, demonstrations and curriculum studies," and also "assists in making exploratory studies designed to determine the feasibility of more extensive research on specific problems."[2]

   Proposals are submitted through the Bureau of Research,[3] U. S. Office of Education, HEW. There are no specific deadlines for small contract proposals, and the program is designed so that proposals are

---

[1] Authorized by the Cooperative Research Act of 1954, with extension through the Elementary and Secondary Education Act of 1965.

[2] College and Univ. Reporter (1651)

[3] Through the regional offices

processed with a minimum of delay. The reader should refer to the Office of Education pamphlet "Support for Research and Related Activities" for details on proposal format.

3. See references above for discussion of large-scale project or program support under the Cooperative Research Act.

B. Research in Educational Media (HEW)

References: See College and University Reporter (C.C., Inc.) at 1653-1654, 8671 and 10,001. Also for PL 89-209, see 9651.

1. Title VII of the National Defense Education Act of 1958 (PL 85-864 as amended) provides support for research and experimentation in more effective utilization of educational media.[4] This includes television, radio, motion picture films, slides, tapes, programmed instructional devices, and other media designed to supplement instruction. Part A of Title VII provides support for research, while Part B includes authorization for dissemination of media (or information on application of new media).

2. This program, also administered by the Bureau of Research, provides for several types of grants -- both large and small, and has wide applicability across subject-matter areas. It is possible to obtain support of $10,000 or less for projects designed to improve or evaluate media applications within a department or specific course (i.e., minimum generalizability). Small grant proposals are submitted to Office of Education regional offices.

C. National Science Foundation (NSF)

References: College and University Reporter, 4568-4573; and 9001-9051.

---

[4] The National Foundation on Arts and Humanities Act of 1965 (PL-89-209) Provides authorization for research on media in arts and humanities instruction.

1. The Social Science Division of the National Science Foundation provides support for basic research, for research in the anthropological sciences, economic sciences, sociological sciences, and research in the history and philosophy of science. Authorization is through the National Science Foundation Act of 1950 (PL 507).

2. Specific programs among those alluded to above require different guidelines, and each program includes individual deadlines for submitting proposals. Be sure to note deadlines in planning a proposal. You should count on several months of thought, writing, and criticism from colleagues in preparing a research proposal. This is extremely important.[5]

3. One recent emphasis within NSF has been on science education, and there are now three divisions within the foundation which are responsible for the various science programs in education: Division of Pre-college Education, Division of Undergraduate Education, and the Graduate Education Division. Each division is responsible for a variety of science education programs, many of which overlap those of the U.S. Office of Education.

D. Arts and Humanities Endowment Funds

References: College and University Reporter, 1720, 1722, 1724, 9651.

1. Research support is available for both broad and specific studies, with emphasis upon American history and literature.

2. The Humanities Endowment cooperates with the Office of Education in sponsoring research on teaching of the humanities at the pre-college level.

3. Send for pamphlet: National Endowment for the Humanities - Initial Programs, September 1966, for additional information, or refer to source listed above.

---

[5] NSF proposal guidelines may be found in College and University Reporter, beginning at 5401.

E. Other Sources*

1. Office of Naval Research (DOD)

2. Department of the Army (DOD)

3. Advanced Research Project Agency (DOD)

4. Agency for International Development

*See Guide to Support Programs for Education, 3M Company, St. Paul, Minn., 1967

*See also pgs. 16 and 17 of the USOE pamphlet (Support for Research and Related Activities) for programs administered by the Bureau of Research.

F. National Institute of Mental Health[6] (NIMH)

**References: College and University Reporter, 4125 through 4128.

1. The NIMH supports basic and clinical research relating to the etiology, diagnosis, treatment, and prevention of mental illness. Grants are also available for instructional research and teaching.

2. NIMH further provides funds in support of research on mental retardation. Support is available on the treatment, care, management and training of mentally retarded in light of biological, psychological, or sociocultural factors involved.

**Future emphasis will be on "research on child rearing practices, teaching programs, parent and child therapy techniques, and interdisciplinary approaches to treatment and rehabilitation."[7]

---

[6] Divsion of the Public Health Service, HEW.

[7] College and University Reporter, 4126.

G. Child Health and Human Development

References: College and University Reporter, 4111

1. The National Institute of Child Health and Human Development - a division of the National Institute of Health (NIH) of the Public Health Service (Department of Health, Education and Welfare) - provides support for research and training related to "maternal health, prenatal care, child health, and human development." NIH emphasizes research in four areas: reproduction, growth and development, mental retardation, and aging.

2. Specific guidelines and notification of deadlines for proposals should be requested from the National Institute of Child Health, Public Health Service.

Appendix E

Utilizing Library Resources in a Search of Literature

The first item to remember when beginning a search through the literature is that the reference librarian is a major source of information. The reference librarian can save countless hours, if only the researcher will present the librarian with information, specific if possible, on what is being sought. The reference librarian can quickly point to the most pertinent information. Crucial books may be kept at the reference desk, so don't overlook asking for suggestions about books which may be held behind the desk. For example, the book How to Locate Educational Information and Data is a reference source which is usually kept at the librarians desk.

The card catalog itself is a major resource. Most large libraries have extensive subjects headings, in red, for subjects pertinent to research. If "educational research" is not located under that heading, try "Research, Educational". Books which contain summaries of educational research may be located through the card catalog. Also, large libraries have special librarians who are authorities in their field, i.e., social science, education, etc. This specialist can be called upon if the reference librarian is not aware of some of the resources.

Indexes: Major sources of information for educational researchers are:

(1) Psychology Abstracts (indexes of authors and subjects at the end of the volume; also separate bolumes containing indexes to authors and subjects, cumulative for the past several years, are now available);

(2) Education Index, classified by author, title and subject (remember that in most indexes author and subject are usually more accurately included than are titles);

(3) Sociology Abstracts (have separate indexes, of author and subject);

(4) Dissertation Abstracts (has a cumulative index for the year);

(5) There are technical indexes such as Operations Research/Management Science, U.S. Government Research and Development reports (prior to 1963 indexing is terrible, almost impossible to find anything except under subject);

(6) Business and Periodical Index, which contains references to articles dealing with business education, for example.

(7) Books in Print, Subject Index; and,

(8) American Library Association, Standard Catalog.

(9) Become acquainted with the E.R.I.C. System. This Educational Research Information Center, established by the U.S.O.E. Bureau of Research, appears to be developing into the most useful and comprehensive source. See in Section X, the Support for Research booklet, p. 5.

X - USOE GUIDES

Section X

USOE Support

Two useful documents from the U.S. Office of Education are contained in this section.

One describes the activities supported by the Bureau of Research. The appendices listing Regional Offices and ERIC Clearinghouses may be of particular value in your initial research effects.

The other describes the Small Project Research Program. This program is highly relevant to the background and interests of participants in this Institute. The proposal format and application procedure are included.

OFFICE OF EDUCATION SUPPORT FOR RESEARCH AND RELATED ACTIVITIES

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
John W. Gardner, Secretary

Office of Education
Harold Howe II, Commissioner

Bureau of Research
R. Louis Bright, Associate Commissioner for Research

FEBRUARY 1967

CONTENTS

APPENDIXES

Foreword

SYSTEMATIC RESEARCH, DEVELOPMENT, AND DISSEMINATION are indispensable to the improvement of education. They are the means by which the educational enterprise assesses its current strengths and shortcomings, works out new programs and techniques to meet changing needs, and provides objective information and models on which the schools and the public can base their individual steps toward educational improvement.

Through support of research and related activities conducted outside the Office of Education, the Bureau of Research encourages efforts to develop and provide the kind of education today's society needs and tomorrow's society has every right to demand. Research support complements other forms of support, whether from local, State, private, or Federal sources. It provides useful blueprints of promising alternatives and thus gives direction and momentum to education's forward thrusts.

This booklet describes the wide range of activities which receive support administered by the Bureau of Research and the general application procedure to be used by those who seek support under any of the Bureau's authorizations.

R. Louis Bright
Associate Commissioner for Research

INTRODUCTION

THE OFFICE OF EDUCATION'S BUREAU OF RESEARCH seeks to improve education through support of a variety of research and related activities. Through authorizations from the Congress, the Bureau provides funds for research projects and programs designed to expand knowledge about the educational process, to develop new and improved educational programs and techniques, to disseminate the results of these efforts to educators and the public, and to train researchers in the field of education. This booklet presents an overview of the kinds of activities administered by the Bureau. Its appendixes contain general application instructions (including criteria for selection of proposals for funding), summaries of pertinent legislative authorizations, and other information of interest to applicants for support.

One major factor distinguishes activities eligible for support through this Bureau: They must be research or research related, as distinguished from operational activities which receive support through other Office of Education bureaus.[1] Research, development, demonstration, and dissemination help provide objective bases for the improvement of operational programs at all levels of American education.

The Bureau of Research awards support under a number of legislative authorizations, the most flexible of which is the Cooperative Research Act, as amended by Title IV, P.L. 89–10. Under this act, it is possible to award grants or contracts for any research or related activities which promise to benefit education. Other authorizations permit support of research and related activities in a variety of categorical areas: Educational research in foreign countries; research and development relating to foreign language teaching and language development; research on educational uses of new communication media and dissemination of the resulting information; research and demonstrations dealing with the education of handicapped children and youth; research, development, and training in vocational and technical education; and library and information science research. (See appendix B for information on legislative authorizations administered by the Bureau.)

Concentration of administrative responsibility for all these authorizations within a single Bureau simplifies application procedures and facilitates coordination among programs. Bureau staff receive all incoming proposals, analyze them to determine whether they are suitable for Bureau support, and allocate them for necessary internal and external review. The Bureau has five divisions: Adult and Vocational Research, Elementary-Secondary Research, Higher Education Research, Research Training and Dissemination, and Laboratories and Research Development. There are also separate staffs for the Arts and Humanities Program and for the Regional Research Program in the Office of the Associate Commissioner. Regional offices are being staffed to administer Small Project Research and Research Development Grants for small and developing colleges. (See appendix C for information on the nine OE regional offices.)

[1] For information about Office of Education support for nonresearch activities, inquiries should be sent to the Bureau of Elementary and Secondary Education, the Bureau of Higher Education, or the Bureau of Adult and Vocational Education, whichever is applicable.

PATTERNS OF SUPPORT

There are two patterns of research support available from the Bureau of Research: Project support and program support.

Application for project support should be made when the initiator wishes to engage in a self-contained activity or set of activities with a few well-defined objectives that can be carried out in a predetermined period of time. A great many different kinds of projects can be supported simultaneously under this pattern. Program support is applicable for specifically announced problem areas in education where the Office of Education feels there is a need for continuous, intensive attention. Programs provide for a concentration of professional resources on these areas over an extended period of time. Thus, project support deals with clearly delineated, limited-time research (of any magnitude), while program support is reserved for certain groups of continuous research or research-related activities which are able to adapt to evolving situations.

Projects are usually planned and initiated by those who submit them as proposals. The relatively flexible format of the General Application Instructions, appendix A, frees the applicant to focus upon the activity or group of activities he proposes to carry out. These instructions should be sufficient for applying for support for most projects. However, those seeking program support may need to request special sets of instructions that supplement the material in this booklet.

Project Support

By far the greatest number of activities administered by the Bureau receive project support. Subjects of project research are as varied as the questions educators seek to answer. For example, some projects may explore educational needs or seek to resolve issues. Others may develop materials or methods, or test them in controlled situations or field studies. Still others may investigate the potential of promising programs or practices to bring about desirable educational change—in fact, they may investigate factors related to the change process itself. Projects may analyze, consolidate, or synthesize information from research or from practice. They may demonstrate or disseminate educational information or techniques.

Some projects may seek to advance knowledge or solve theoretical problems in the behavioral sciences. Others may involve applied research or curriculum development in specific content areas or combinations of areas. Still others may produce materials for educational technology or demonstrate their uses. Projects may be directed toward any educational level or toward students with particular problems or characteristics. In short, the Bureau adjusts its support to whatever kinds of projects promise effective and practicable innovations. In this way, it provides freedom for a viable, self-renewing approach to educational improvement.

Project Research in General.—There is no limitation on the size, the area of study, or the kind of activity or activities eligible for project support—so long as they deal with educational research, development, or dissemination. However, the public interest and the critical needs of education demand careful administration of available funds. Sometimes this means judicious selection from among similar proposals to avoid unnecessary duplication. Sometimes it means making support of a promising research activity contingent upon the investigator's ability to redesign his research so that its results will be useful in settings other than his own. The size of a study is no measure of its potential for educational improvement. In some cases, a study must be quite large in order to produce valid information; in other cases, a relatively inexpensive pilot study may accomplish a great deal. When it is necessary in terms of total research goals, the Bureau may solicit proposals in critical areas where research interest has been slow to develop.

All proposals are assessed in terms of their promise for meeting stated objectives, the significance of those objectives to the Office of Education's *total* effort, and economic efficiency. Small-scale research or development projects can be

funded with a minimum of delay for proposal review. Consideration of larger projects takes longer.

Small Project Research refers to those activities which require no more than $10,000 from the Office of Education and take no longer than 18 months for completion. This special classification has been provided to give adequate consideration to inexpensive yet worthwhile projects; to encourage personnel of small colleges to gain experience in research and related activities; and to support significant small-scale educational research projects by doctoral and postdoctoral students and fellows, particularly those at developing institutions.

Support for these activities is being decentralized to the OE regional offices. This method of administration should facilitate prompt consideration of proposals and bring evaluation, negotiation, monitoring, and other assistance closer to those carrying out the projects. The instructions in appendix A are generally applicable to small projects, but separate instructions are available by writing to the regional offices or to the Regional Research Program staff in the Bureau.

Program Support

In addition to the wide array of projects described above, the Bureau of Research also supports continuous programs, each of which provides a particular thrust in the total research and development effort. This type of support is used for the activities of research and development centers and educational laboratories, for training educational researchers, and in other appropriate circumstances. These programs allow long-term staff commitments and continuous development and adaptability. Appropriate supplementary instructions for applying for program support are available from the Bureau of Research. (See programs below.)

Research Development Grants support the efforts of small or developing colleges to acquire sound research orientations. Their personnel may use the funds to develop their own research skills, to research their own educational problems, or to try out promising educational innovations. A primary purpose is to help teachers and future teachers learn to use research, research results, and the research or inquiry approach in their classrooms.

To encourage the development of research capabilities in smaller institutions of higher education, especially those which train teachers, grants are available to consortia or groups of institutions which plan to combine their ideas and competencies in a research development activity. Single institutions may also apply but should scale their programs and fund requests appropriately. Participating institutions should expect to (1) disseminate information on research findings and research design, administration, and management; and (2) provide their staffs with time and money to conduct the research activities.

OE regional offices are being staffed to administer Research Development Grants. Separate instructions with examples of appropriate activities are available either from the Bureau of Research or from the regional offices.

Research and Development (R & D) Centers.—A research and development center concentrates on a single problem area in education and conducts activities ranging from basic research through dissemination. Centers are usually established at universities or other institutions where staff have already demonstrated exceptional competency in the particular problem area and can be expected to produce early, continuous, and significant educational advances. The sponsoring institution generally continues to provide s[illegible]antial local support for center activities.

Each center is interdisciplinary and ordinarily maintains cooperative relationships with regional laboratories, State departments of education, local school systems, universities and teacher training colleges, and relevant professional and nonprofit organizations. Within its established area of investigation, each center can direct its own program without obtaining prior Office approval for individual projects. Thus, center ac-

258-410 O - 67 - 2

tivities can reinforce each other, promising leads from one activity can be immediately followed up, and research findings can quickly be put into practice.

At the end of fiscal 1966, 11 of these centers were in operation, each working in a major problem area. The comprehensiveness of center operations and the cost of staff dictate extremely careful evaluation of proposals for establishment of new centers. Care must be taken to select only problem areas which need continuing investigation. Each center must have the professional and administrative competencies to direct activities which constantly evolve from these investigations.

Those persons seriously planning to apply for support to establish an R & D center should consult with Bureau staff of the R & D Center Program and secure supplemental instructions before drawing up a proposal. These instructions explain the special conditions applicable to support for R & D centers, including site visits, monitoring, and arrangements for continuation funding. Individuals interested in working with one of the centers should apply directly to the center, and not to the Bureau of Research. A list of R & D centers, their locations, and their areas of inquiry is provided in appendix D.

Educational Laboratories.—Educational laboratories differ from R & D centers in their focus, composition, and activities. Although both may work all along the continuum from basic research to dissemination and implementation, centers emphasize research and development while laboratories stress development, dissemination, and implementation. Each laboratory is primarily concerned with educational improvement in a particular region, especially with wider adoption of beneficial educational innovations there. To this end, the laboratory designs its own program and continuously adjusts it to meet emerging needs of the region. In terms of organizational structure, the laboratories are new institutions which draw upon colleges, universities, State educational agencies, local schools, private industry, and other educational interests for their staff, membership, and affiliations.

Laboratories are set up through the initiative of individuals and groups in the regions. These local and regional efforts have resulted in 20 laboratories, generally organized as nonprofit corporations, which form a network serving all of the continental United States. Individuals and groups interested in effecting educational improvements through the regional laboratory program should contact the laboratories directly, not the Bureau of Research. A list of the 20 educational laboratories with their addresses and participating States is provided in appendix E.

Programs for Training Educational Researchers.—To provide for sound educational research and development in the future, support is available for institutions to train researchers and to develop and improve their own programs for such training. Institutions may request support for undergraduate, graduate, and postdoctoral training; and for training institutes, inservice programs, or special projects dealing with educational research. Funds may be used to develop and strengthen research training staffs and curricular capabilities and for stipends and institutional allowances for trainees. Training may be concerned not only with research per se but also with educational strategies needed to bridge the gap between research and practice.

Ability to provide such training, potential for program development, and equitable geographic distribution are considered in the establishment of these programs. Cooperative programs among sponsoring institutions and with State agencies, regional laboratories, and other educational groups are encouraged. Special instructions and application forms are available from the Bureau's Division of Research Training and Dissemination. Individuals wishing to participate in these research training programs should apply to the institutions which provide the programs, not to the Bureau of Research.

Through a limited number of National Postdoctoral Fellowships in Educational Research, the Office of Education arranges with institutions to provide a year of intensive research experience

for selected candidates. Further information on Postdoctoral Fellowships is available from the Bureau's Division of Research Training and Dissemination.

Educational Research Information Center (ERIC). —ERIC is a comprehensive national information system designed to serve American education by making available reliable, current educational research and research-related materials. The system is made up of a network of information clearinghouses or documentation centers located throughout the country and coordinated through Central ERIC in the Office of Education.

By the end of 1966, clearinghouses had been established in 13 substantive areas (see appendix F). As funds become available, the Bureau will issue requests for proposals to establish additional clearinghouses. Each clearinghouse collects materials in a different subject area and is staffed by specialists who are responsible for document analysis, selection, and other activities related to ERIC's mission. Documents collected throughout the ERIC system are abstracted, indexed, and put on microfiche, 4" x 6" film cards that contain up to 60 pages of text per card.

Starting in November 1966, a monthly journal, RESEARCH IN EDUCATION, is being issued by the Office of Education through the Bureau of Research, Division of Research Training and Dissemination.[2] Each issue contains bibliographic citations and abstracts of recently funded projects and of final reports of completed projects supported through the Bureau of Research, as well as detailed indexes of cited research documents. Subsequent issues will include documents from the 13 ERIC clearinghouses. A cumulative index will be published annually.

Other Specialized Programs.—In addition to the relatively general programs described above, the Bureau of Research also administers support for the following specialized programs.

Training of Personnel in Vocational Education may be provided through Bureau-supported institutes to prepare teachers for new and changing occupations and to upgrade the competencies of teachers already engaged in adult and vocational programs. Participants may include not only vocational teachers and counselors but also school administrators and related vocational-technical-adult education personnel. Support may also be requested for seminars to improve competencies of individuals concerned with vocational education research and development activities. Forms and instructions for making application are available from the Division of Adult and Vocational Research, Bureau of Research.

State Research Coordinating Units in Vocational Education, located in all but a few States, coordinate local research and demonstration efforts in vocational education and provide consultative assistance to individuals who wish to prepare research proposals in that field. Support administered through the Division of Adult and Vocational Research is gradually being reduced as States assume this responsibility.

Instructional Material Centers for Handicapped Children and Youth are being established to provide a network of repositories for such aids as braille books, test kits, and tapes and recording devices where staff can evaluate their effectiveness and make them available to local schools. The centers will also engage in research and development aimed at improving teaching materials for the handicapped and hold institutes and workshops to familiarize teachers with the use of special educational materials. Appendix G lists the 10 centers established by the middle of 1966. As funds become available, new centers may be added to serve particular geographic areas. Information is available from the Handicapped Children and Youth Research Branch, Division of Elementary-Secondary Research, Bureau of Research.

---

[2] Subscriptions ($11 domestic, $13.75 foreign, $1 single copy, domestic) are available from the Superintendent of Documents, U.S. Government Printing Office, Washington D.C. 20402. All reports cited in RESEARCH IN EDUCATION are available as hard copy or microfiche from the ERIC Document Reproduction Service, operated by Bell and Howell Company, 1700 Shaw Avenue, Cleveland, Ohio 44112, under an OE contract. Prices are quoted as part of each citation.

APPLICATION PROCEDURES

The Official Application

Application for support is made by submitting a standard *formal* proposal, accompanied by an official application form. This proposal is evaluated for its merit by the Office of Education staff and by non-Federal field readers selected for their research or specialized experience and general knowledge of the field. It is the only contact between the reviewers and the initiator's idea; if it does not convey the message, the staff and readers will not assume meaning or intent. If the activity is approved for funding, the proposal document becomes part of the contract.

It is suggested that the initiator ask several persons who are not close to the problem to read the proposal to make certain that it communicates clearly. Many potentially good research proposals fall short of recommendation for negotiation because the procedures are not clearly written or omit essential details. Others present so many questions for research that the study would be unwieldy or could not be completed with the funds and within the time requested. Should the proposal fail to be selected for funding, a summary of the reviewers' comments may be obtained upon request and a revised proposal may be submitted. However, reconsideration by the Bureau is no guarantee of subsequent selection.

Preliminary Review

Before preparing a formal proposal for either project or program support, an individual or group may find it helpful to solicit suggestions from the Bureau by sending three copies of a brief (not over five pages) description of the activities to be undertaken to determine whether they would be eligible or appropriate for consideration under Bureau authorizations. The description should include a title, a list of objectives or questions which the research would answer, a brief description of the procedures, and an indication of the significance of the activity for educational improvement beyond the setting in which it would be carried out. The first page should be clearly marked "Preliminary Statement" and should include the applicant's full name, his professional title and the organization with which he is affiliated, and his telephone number, including the area code. The preliminary statement should be sent to the mailing address for formal proposals (see page 12).

This preliminary review is intended to help the potential investigator avoid spending time on a formal proposal that would be ineligible or inappropriate. However, suggestions from the Bureau should not be interpreted as either denying the proposer the right to submit a formal proposal or as guaranteeing approval. Since Bureau staff are obliged first to give attention to review of formal proposals and to monitoring of approved projects, preliminary drafts are reviewed as time permits. The potential applicant, therefore, is encouraged to forward his prospectus well in advance of the time he plans to submit a formal proposal.

APPENDIXES

APPENDIX A. GENERAL APPLICATION INSTRUCTIONS

Activities Appropriate for Support

To be eligible for support administered by the Bureau of Research, an activity must (1) be research or research-related, (2) show promise of improving education, (3) have general (not purely local) applicability, and (4) be directed toward communicable results.

The term "research related" is broadly interpreted to include such activities as the development of materials and improvement of instructional practices in general and specific areas, and dissemination and implementation of the results of research. Local projects must lead to findings significant for other settings if they are to be considered eligible for support under research authorizations. This applicability or transferability to other settings is the factor which, in many instances, determines whether a given proposal is eligible for consideration by the Bureau of Research.

The Bureau's authorizations do not include funds for general operating programs or for activities carried on in conjunction with the research which are part of the normal activities of project personnel (e.g., teacher salaries in experimental classes). Bureau support may be requested for the research component of larger projects where the operating costs are funded from other sources. In a few instances, support may be provided for those aspects of an operational program which must be modified or established in the interests of the project.

With the exception of Small Project Research and Research Development Grant activities, there are no established time or cost limitations for projects which may be considered for support. However, contractual arrangements usually obligate project funds only out of the current year's appropriations. If a proposed activity involves extensive costs or must be continued over subsequent years, the proposal must undergo substantial additional evaluation to assure sound initial investment.

Eligible Parties

A record of noncompliance, by an applicant, with the terms of prior research grants or contracts (for example, failure to submit a final report) may be sufficient grounds to refuse consideration of a proposal.

Grants or contracts may be awarded to colleges, universities, State departments of education, or to other public or private agencies, organizations, groups, or individuals after proposals have been reviewed by Office of Education staff and by appropriate non-Government advisory personnel and approved by the U.S. Commissioner of Education. With profit-making organizations, contracts follow Federal Procurement Regulations.

The type of support instrument—i.e., grant, cost-reimbursable contract, or fixed price contract—may be selected by contract personnel during negotiation. If the authorization providing the funds does not specify which type of support, consideration is given to providing the best balance between fiscal control and flexibility of operation. All research grants must include significant cost sharing by the grantee institution. Moreover, in some contract research activities, non-Federal contributions are appropriate and a factor in award consideration.

Eligibility for larger projects and for the beginning phase of longer projects may be limited to groups which are or can be organized so as to ensure proper continued professional and fiscal accountability and competence.

Public schools, school districts, and local educational agencies, in order to assure appropriate liaison and communication with operating school systems, are requested to send three copies of their proposals to their State department of education for comment. To be eligible for support under the Vocational Education Act of 1963, proposals submitted by local educational agencies must contain evidence of approval by an authorized official of the State board of vocational education.

Criteria for Evaluation of Proposals

Proposals are evaluated and recommended for approval according to these criteria:

1. Educational significance.
2. Soundness of design, procedure, or plan.
3. Adequacy of personnel and facilities.
4. Economic efficiency.
5. Other specific criteria, as appropriate.

Evaluation of a proposal's significance to education requires more than consideration of the project itself. It involves attention to the breadth of the project's probable impact, its relationship to other ongoing and completed research, and its capacity for contributing to educational improvement within the context of total research needs.

Educational significance, then, is the first of several conditions for support; but if a proposal lacks adequate technical quality, personnel, or economic efficiency, it will not be supported, regardless of its significance. In like manner, no matter how technically excellent the proposal may be, its support is contingent on its significance to education nationally.

Funds are not available to support all the good proposals submitted to the Bureau of Research; the current approval rate is about one in five. To provide a balanced effort toward educational improvement within the context of active competition for available Federal research funds, it is sometimes necessary to forego support in a well-researched area in order to attend to a neglected but critical one. There also are instances when difficult choices must be made between one or two relatively large projects and a larger number of smaller ones. In any case, selection of any given activity for support is based on systematic evaluation of the plan set forth in the formal proposal document and determination of whether it meets the needs of education.

The Proposal Document

Application for support is made by submitting a properly executed standard application form (available from the Bureau of Research or regional offices) and the appropriate number of copies of a formal proposal which describes the activity or activities, explains their significance, identifies key personnel, and estimates costs.

Since funds are available only for the most promising activities, proposals must be assessed competitively (see Criteria for Evaluation of Proposals above). A uniform proposal format has been designed to accommodate all Bureau of Research projects, regardless of the authorizations under which they are funded. Use of a single format gives the applicant freedom to concentrate on the particular activity to be undertaken without having to choose among application patterns.

All proposals must include the standard cover page, a one-page abstract, the body section, and personnel and budget items. Within this framework, each applicant states the case for his activity.

The guidelines below are generally applicable, regardless of the magnitude of the proposed project, the area of investigation, or the authorization under which the project may be funded, if approved. The applicant is expected to make judicious adaptations of this format to accommodate the kind of research or related activity he proposes to undertake.

I. *The Cover Page.*—Nothing may precede this page. It contains only the information on page 10 in the order indicated, and two copies must be signed by the Initiator or Project Director and by the Transmitter.

II. *The Abstract* occupies a single page, identifies the proposal, and concisely and simply summarizes the contents. To accommodate the various uses made of this page, the abstract must be written in language understandable by an informed layman. One abstract is placed after

PROPOSAL FOR RESEARCH AND/OR RELATED ACTIVITIES
SUBMITTED TO THE U.S. COMMISSIONER OF EDUCATION FOR
SUPPORT THROUGH AUTHORIZATION OF THE BUREAU OF RESEARCH

Title: (Be concise; avoid obscure technical terms)

Applicant Organization: (Name of College or University, State Agency, School District, or other unit with major commitment to the activity)*

Initiator:
(Principal Investigator)

Signature

(Full name of person responsible for development of the proposal; position; telephone area code, number, extension)

(If initiated by a student under faculty sponsorship, add line—sponsored by

Signature ,—and give the same information for the sponsor.)

Project Director:
(if different from Initiator)

Signature

(Full name of person in charge of investigation; position; telephone area code, number, extension)

Transmitted By:

Signature

(Full name and position of official committing the applicant organization or cooperating institution, agency, school district, etc., to the activity; telephone area code, number, extension)

Duration of Activity: (Proposed beginning and ending dates)

Total Federal Funds Requested:

Date Transmitted:

* Individuals or private organizations should adapt the format of this cover page to show information comparable to that required of others.

the title page in each copy of the formal proposal, and 20 copies are provided separately. At the top of the abstract page, the following items should appear on separate lines: Title of Project, Principal Investigator, Contracting Agency, Amount of Federal Funds Requested, and Proposed Beginning and Ending Dates. The summary portion of the abstract has three parts: A statement of the purposes, objectives, or nature of the project; an indication of the expected contribution to education; and a compendium of procedures or description of what is to be done.

III. *The Body* of the proposal communicates the project director's plan and its probable effectiveness. It should be clear, concise, forthright, and complete, and should be kept within 20 single-spaced, typewritten pages. Proposals for exceptionally extensive or complicated projects may subsequently be developed into larger documents with the advice and cooperation of Office staff. The body of Small Project Research proposals should not exceed 10 pages.

Within the body section, there are normally three parts which should be subtitled as appropriate for the particular kind of research, development, or dissemination activity to be undertaken. These parts answer the questions: Why? What, when, and how? and Use of results? Some attention is also given to personnel, facilities, and costs, which are subsequently treated in detail.

A. *Problem and Objectives.*—The first part tells *why* the research or related activity should be undertaken. It includes a statement of the problem or purposes, review of literature and related research, concise statement of objectives, or any other information necessary to establish a sound rationale.

B. *Description of Activities (Procedures).*—This part tells *what* is to be done, *when*, and *how*. It is the basis for determining the degree to which the proposed activity can be expected to accomplish the objectives or satisfy the need set forth in the first part of the proposal body. It delineates procedures, outlines program arrangements, describes materials to be produced, or otherwise explains how the activity leads to results and how evaluation will be accomplished. Allowances for alternatives, if any, should be noted. (A time schedule for completion of the project is usually provided near the end of this part of the proposal.)

The amount and kind of detailed information will vary according to the type of activity. For example, a basic research study—which might include rather strict statistical treatment—would be approached quite differently from a curriculum development activity. A project which is complete within itself would not be approached in the same way as a phase of a proposed multiphase activity. Preparation of initial materials for a modern foreign language would not include the same procedures as the demonstration of new media usage or the testing of a pilot study in vocational or technical education.

C. *The Use To Be Made of Findings.*—This part tells how the results of the activity may be disseminated and/or implemented, what contribution to education can be expected, and what steps should follow.

IV. *Personnel and Facilities.*—Personnel and facilities are a major determinant of capability.

Personnel with major responsibilities are listed by name, position, title, experience, responsibilities within the project, percentage of time committed to the activity, and the extent to which this commitment has been assured. Consultants who have agreed to serve should be similarly identified (otherwise, the application should describe the type of consultative assistance required).

Facilities should be described, and the extent to which their use has been assured should be indicated.

V. *The Budget Section* of the proposal starts on a new page and uses the tabular presentation shown at the end of appendix A. The applicant should show reasonable estimates but be detailed enough to suggest careful analysis of expected costs and understanding of fiscal responsibilities

in connection with conducting the proposed activity. Alternative budgets may show allowances for extension of objectives or revision of procedures. Local contribution, or cost sharing, is *required* on all grants, *encouraged* on contracts. Cost sharing may not be avoided by substitution of a contract in a situation requiring a grant instrument. Local contribution, while not necessarily a measure of the project's soundness or potential worth and not specifically required on certain contracts, is considered evidence of local interest and commitment in the activity. The budget section should include an outline of the sources and amounts of non-Federal support and any conditions upon which this support is contingent.

VI. *Appended Items*

Other Information.—Give a brief statement about each of the following: (a) If this or a similar proposal has been submitted elsewhere, give details. (b) If this is a proposed extension of, or addition to, a previous or current project supported by the Office of Education, give the Bureau and the grant or contract number of the related support instrument. (c) If there has been any previous communication with the Bureau of Research on this proposal, give name of staff member concerned.

Revisions.—If this proposal is a *resubmission* of a previous formal proposal, give the Bureau proposal number assigned the original proposal and describe the major revisions which have been made. (This does not apply to preliminary statements submitted for informal review.)

Report of Other Projects.—If any of the primary personnel have a *current* or *uncompleted* project with the Office of Education or other agency, an appended statement should indicate the status of the project, the amount of time devoted to it, and the relationships between the current and the proposed project.

If any of the personnel have *completed* a research or development project supported by the Office of Education, give information to identify it. If findings of the previous project are related to the current proposal, summarize them briefly.

Agreement with Cooperating Agencies.—Where agreements with school districts or other cooperating agencies are a factor, copies should be appended.

Instruments.—When an applicant plans to use a published or unpublished instrument (e.g., a questionnaire or interview guide) in his study, the proposal document should include a copy or, if the instrument is still to be developed, a page of sample items and an outline of the complete instrument. (See Special Conditions, Data-Collection Instruments on page 14.)

Other Items.—If necessary, other items may be added, but appended items cannot carry the burden of the request for support.

Details About Submitting Proposals

Forms for the standard application should be requested from the Bureau of Research at least 2 weeks prior to the date the copies of the formal proposal are to be submitted.

Proposals should be on one side of standard (8½" x 11") paper, stapled at the left margin. Do not bind or enclose in folders. Cover pages on two copies should be signed by the Initiator or Project Director and by the Transmitter.

Address

The address label for most proposals should read:

Research Analysis and Allocation Staff
Bureau of Research
U.S. Office of Education
Washington, D.C. 20202

Note.—If the OE regional office in the applicant's area has been staffed with an Educational Research Adviser, proposals for Small Project Research and Research Development Grants should be sent there. Otherwise, they should be mailed to the above address. When in doubt, contact the Bureau or the appropriate regional office listed in appendix C.

Number of Copies

Small Projects.—Send 20 copies of proposals requesting $10,000 or less in Federal support, with appropriate copies of abstracts.

General Projects.—Send 30 copies of larger proposals. The Bureau may request 10 additional copies of complex proposals. Besides the abstract inside the cover sheet of each proposal, 20 copies of the abstract should be provided separately.

Review Procedures

There are no deadlines for general submission, but some of the special programs process proposals only at certain times. Proposals are first reviewed by Office of Education staff. Those which are within the scope of the Bureau's legislative authority and show promise of contributing to educational improvement are then reviewed by specialists outside the Government. These groups make recommendations to the Commissioner as a basis for his decision and final approval. While many proposals may be processed within 3 or 4 months (Small Projects within 2 months), larger or more complicated ones may require more time.

Communications Regarding Proposals

The initiator will be sent the name of the Bureau of Research unit, the project officer, and the official identification number assigned to his proposal. The Office does not give further information about the status of the proposal during the review and recommendation process. When the proposal has been thoroughly assessed in terms of the criteria listed earlier in this booklet and within the context of pressure for available research funds, the initiator will be notified whether his proposed activity is to be recommended for negotiation. If the project is so recommended, the initiator will be sent any information necessary for subsequent action.

Contractual Procedures

Negotiation is that process by which a legal and mutually satisfactory agreement (contract or grant) is arranged between the applicant and the Office of Education. This process transcends a mere final examination of estimated costs. The contract specialist works with the research specialists (applicants and appropriate staff from the Bureau of Research) to clarify and set forth their mutual agreements about the work and services to be carried out and the support to be provided. The resulting contract or grant instrument, which incorporates the final document and/or plan of operation, thus states the conditions under which the award is made by the Office.

The investigator is cautioned that no costs should be incurred prior to the receipt of a contractual document (grant, contract, or notice of award) executed by the Contracting Officer on behalf of the Government. All contracts are subject to satisfactory fiscal and progress reports. In those cases where only an initial phase of a total research project has been supported, continuation is subject to reappraisal of project aims and accomplishments in relation to emerging research needs and availability of funds in any subsequent period.

Reports

Reports to be received by project officers include: (1) generally brief monthly and/or quarterly progress reports which are intended to facilitate project monitoring; (2) two types of substantive reports, (a) interim reports as appropriate and (b) final reports[3]; and (3) appropriate fiscal reports.

Interim reports include any written material (textual, graphic, or tabular) prepared during the project period and generally before submission of the final report. Often such reports are prepared for delivery at a professional meeting, for distribution to colleagues, or for some other kind of presentation of substantive findings. Regardless of the number or variety of progress and interim reports submitted, the project director also must submit 15 copies of a comprehensive final report.

[3] Detailed instructions for preparing these interim and final reports and related information storage and retrieval items are provided at the time the project is funded.

For projects that will produce nonprint materials, the project officer should specify at the time of project negotiation the number of copies of such materials to be received by the Office of Education, taking into account loan uses by program staff, possible uses in regional laboratories, and other needs.

Special Conditions

Release of Materials

Materials produced under research contracts or grants will be placed in the public domain. Whenever any materials not first produced in the performance of a particular contract are to be incorporated in the new materials, the contract must include appropriate arrangements for use of materials so incorporated.

Data-Collection Instruments

Prior Approval.—Clearance prior to administration is required for all tests, questionnaires, inventories, interview schedules or guides, rating scales, and survey plans used to collect information on identical items from 10 or more individuals or organizations. This regulation flows from the provisions of Federal law, from technical requirements for proper project-monitoring, and from sound considerations of policy in the administration of public funds for educational research.

Description.—All projects involving data-gathering instruments require submission of a statement giving: The exact title of the instrument; the purpose(s) of the study and the relation of the data-gathering instruments thereto; the nature and size of the sample, including the method(s) of sampling; the locale of the study; provisions for anonymity and confidentiality of response; a brief indication of the nature and extent of statistical analysis of the data; the estimated average time required of the respondent; and a single figure giving the estimated cost of the survey-proper (as distinguished from the total project). Clearance will take all these factors into account; thus, questionnaire items that are permissible or desirable in one sample might not be so in another.

Clearance.—The clearance process requires *six copies* of each instrument, with these exceptions: Copies need not be submitted of instruments which deal solely with cognitive functions or technical proficiency (e.g., scholastic aptitude, school achievement, vocational proficiency); routine demographic information (e.g., age, sex, race, residence, school attendance); or routine institutional information. However, the information specified in the preceding paragraph is required for all instruments, whether copies of the instrument are submitted or not.

In the process of clearance, attention will be given especially but not exclusively to such matters as the unnecessary or offensive intrusion of inquiries regarding religion, sex, and politics; the extraction of self-demeaning or self-incriminating disclosures; and the apparent countenancing of antisocial or immoral behavior. When the respondents include students below college age, the assurance of informed parental consent will be required if unduly sensitive questions are involved.

Those who will need to clear data-gathering instruments should request separate, more detailed instructions from the Bureau.

Treatment of Animals

If animals are involved in research activities supported by the Bureau of Research, their care and treatment must conform to the principles set forth by the Institute of Laboratory Animal Resources of the National Academy of Sciences—National Research Council.

Research Outside the United States

There are special requirements applicable to research activities outside the United States, including Department of State approval.

BUDGET WORKSHEET GUIDE FOR RESEARCH ACTIVITIES

Initiator ______________________ Duration of proposed activity:

Institution or agency ______________________ Beginning date: ________ Ending date: ________

Category	Project cost by line item*	
	Federal	Local
I. DIRECT COSTS A. Personnel (List all position titles such as project director, research assistant, secretary, consultants,** etc. State percent of time on the project, per annum salary, and beginning and ending dates of employment for each person.)		
B. *Employee Benefits* (Summarize benefits such as Social Security, retirement, etc.)		
C. *Travel* (Indicate fares and/or mileage at allowable rate; number of days per diem,** and rate.)		
D. *Supplies and Materials* 1. Project Materials 2. Office Supplies		
E. *Communications*		
F. *Services* 1. Duplicating and Reproduction 2. Statistical (Itemize costs) 3. Testing (Itemize costs) 4. Other		
G. *Final Report Costs*		
H. *Equipment* (Generally, capital equipment may not be purchased with research funds. Certain types of equipment may be available from Government sources as determined by the Office. If not available, cost of rent vs. purchase will be considered.)		
I. *Other Direct Costs* (List)		
J. *Subtotal, Direct Costs*		
II. INDIRECT COSTS (Give basis on which local overhead is computed, who established the rate, when, and period covered.)		
III. TOTAL COSTS		
IV. COST SHARING (Percentage distribution of proposed Federal and local support.)		

* For projects longer than 18 months in duration, use one column for each 12 months of activity and a final column for totals.

** Consultants: Show rate and number of days under Personnel, transportation and per diem under Travel.

APPENDIX B. LEGISLATIVE AUTHORITY FOR RESEARCH AND RELATED ACTIVITIES SUPPORTED BY THE BUREAU OF RESEARCH

(in order of enactment)

P.L. 83-480 Agricultural Trade Development and Assistance Act of 1954 (as amended)
Sec. 104(k), *Research in Foreign Countries.*

The Foreign Currency Program, authorized under the above law, uses a portion of the Nation's foreign currencies for financing a wide range of educational research and related activities carried out in certain designated countries abroad. Almost all of these funds are used under agreements with research organizations in the foreign countries themselves, but some may be used by U.S. domestic applicants. Countries with P.L. 480 funds in 1966 were: Burma, Ceylon, India, Israel, Pakistan, Poland, United Arab Republic, Yugoslavia, Indonesia.

P.L. 83-531 Cooperative Research Act (as amended)
Sec. 2 (a), *Cooperative Research Program*
Sec. 2 (b), *Educational Research Training Programs*
Sec. 4, *Educational Research Facilities*

Inauguration of the Cooperative Research Program in fiscal year 1957 launched the first of the systematic research programs to be conducted outside the Office of Education with Federal support administered by the Office. The 1954 Cooperative Research Act authorized the Office to enter into jointly financed cooperative arrangements with universities and colleges and State educational agencies for conducting research, surveys, and demonstrations in the field of education. The scope of the program was expanded by Title IV of the Elementary and Secondary Education Act of 1965 which added support for dissemination, construction and operation of facilities for research and related activities, and development of programs to train educational researchers. Title IV also expanded eligibility for participation. This program permits the Office to support a variety of educational research projects and programs, thus rounding out a comprehensive research effort.

P.L. 85-864 National Defense Education Act 1958 (as amended)
Title VI, *Language Development*
Sec. 602, *Research and Studies*
Title VII, *Research and Experimentation in More Effective Utilization of Television, Radio, Motion Pictures, and Related Media for Educational Purposes*
Part A, *Research and Experimentation*
Part B, *Dissemination of Information on New Educational Media*

Language Development Research and Studies are authorized by Title VI of the National Defense Education Act of 1958, which provides support for: studies and surveys to determine the need for increased or improved instruction in modern foreign languages and other fields necessary for a full understanding of the areas, regions, or countries where the languages are commonly used; research on more effective methods of teaching such languages and such other fields; and development of specialized materials for use in training students and teachers in these languages or fields.

Media Research and Dissemination, authorized by Title VII A and B of the National Defense Education Act, provides support for (A) research and experimentation in the development and evalu-

ation of activities involving television, radio, motion pictures, printed and published materials, and related media of communication for educational purposes, and (B) the dissemination of information concerning new educational media through studies and surveys, demonstrations, and publications and reports.

P.L. 88-164 Mental Retardation Facilities and Community Mental Health Centers Construction Act of 1963 (as amended)
Sec. 302, *Research and Demonstration Projects in the Education of Handicapped Children*

The authorization for Education of Handicapped Children Research and Demonstration provides support for activities relating to education for the mentally retarded, hard of hearing, deaf, speech impaired, visually handicapped, seriously emotionally disturbed, crippled, or other health-impaired children who, by reason thereof, require special education.

P.L. 88-210 Vocational Education Act of 1963
Sec. 4(c), *Vocational and Technical Education Research, Development, Training*

This provision for Vocational and Technical Education Research, Development, and Training provides assistance in paying the costs of research and training activities and of experimental or developmental programs and projects to help meet the special vocational education needs of youth (particularly youth in economically depressed communities who have academic, socioeconomic, or other handicaps that prevent them from succeeding in regular vocational education programs).

P.L. 89-329 Higher Education Act of 1965
Title II, *Part B, Library Research and Development*

The authorization of Library Research and Development provides support for research and demonstration projects to improve libraries and training in librarianship; to develop new techniques, systems, and equipment for handling and distributing information; and to disseminate information derived from such research and demonstrations.

APPENDIX C. OE REGIONAL OFFICES

Regional Office of Education staff are located at the following addresses:

Region	*States Served*	*Address*
Region I	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont	Office of Education/DHEW John Fitzgerald Kennedy Federal Building Boston, Massachusetts 02203
Region II	Delaware, New Jersey, New York, Pennsylvania	Office of Education/DHEW 42 Broadway New York, New York 10004
Region III	Kentucky, Maryland, North Carolina, Puerto Rico, Virginia, Virgin Islands, West Virginia, District of Columbia	Office of Education/DHEW 220 Seventh Street, N.E. Charlottesville, Virginia 22901
Region IV	Alabama, Florida, Georgia, Mississippi, South Carolina, Tennessee	Office of Education/DHEW 50 Seventh Street, N.E. Atlanta, Georgia 30323
Region V	Illinois, Indiana, Michigan, Ohio, Wisconsin	Office of Education/DHEW New Post Office Building 433 West Van Buren Street Chicago, Illinois 60607
Region VI	Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota	Office of Education/DHEW 601 East 12th Street Kansas City, Missouri 64106
Region VII	Arkansas, Louisiana, New Mexico, Oklahoma, Texas	Office of Education/DHEW 1114 Commerce Street Dallas, Texas 75222
Region VIII	Colorado, Idaho, Montana, Utah, Wyoming	Office of Education/DHEW Federal Office Building, Rm. 9017 19th and Stout Streets Denver, Colorado 80202
Region IX	Alaska, Arizona, California, Guam, Hawaii, Nevada, Oregon, Washington, American Samoa	Office of Education/DHEW Federal Office Building, Rm. 232 50 Fulton Street San Francisco, California 94102

APPENDIX D. RESEARCH AND DEVELOPMENT CENTERS

The twelve R & D Centers in operation as of November 1966 are listed below by date of establishment. Ten are supported by Cooperative Research Program funds and two by provisions of the Vocational Education Act.

Fiscal year established	*Name of center, location, and area of inquiry*
1964	Learning Research and Development Center, University of Pittsburgh, Pennsylvania (Learning research and instructional practices)
	Center for the Advanced Study of Educational Administration, University of Oregon, Eugene (School organization and administration in the societal context)
1965	Center for Research and Development for Learning and Re-Education, University of Wisconsin, Madison (Learning efficiency for children and adults)
	Center for Research and Development on Educational Differences, Harvard University, Cambridge, Massachusetts (Effects of individual and cultural differences on the learning process)
	Center for Research and Leadership Development in Vocational and Technical Education, Ohio State University, Columbus (Research and development activities, including operation of ERIC clearinghouse on adult and vocational education)
	Center for Research, Development, and Training in Occupational Education, North Carolina State University, Raleigh (Research and development emphasizing southern needs in adult and vocational education)
1966	Research and Development Center in Educational Stimulation, University of Georgia, Athens (Programs of early and continuous stimulation, 3- to 12-year-olds)
	Research and Development Center in Teacher Education, University of Texas, Austin (Teacher education)
	Stanford Center for Research and Development in Teaching, Stanford University, Palo Alto, California (Theory and practice of teaching and its effects)
	Center for Research and Development in Higher Education, University of California, Berkeley (Organization, purposes, and outcomes of higher education)
	Center for the Study of the Evaluation of Instructional Programs, University of California, Los Angeles (Study of evaluation processes and techniques)
1967	Center for the Study of Social Organization of Schools and the Learning Process, Johns Hopkins University, Baltimore, Maryland (Influence of social and administrative organization of schools on learning of students from diverse backgrounds)

APPENDIX E. REGIONAL EDUCATIONAL LABORATORIES AND PARTICIPATING STATES

As of November 1966 the following Regional Educational Laboratories had been established to serve every section of the continental United States, and a feasibility contract had been negotiated to study needs and resources for a laboratory in the Hawaii-Pacific Basin area.

Appalachia Regional Educational Laboratory, Charleston, West Virginia (West Virginia, the Appalachian counties of Virginia, Tennessee, Kentucky, Ohio, and Pennsylvania)

Center for Urban Education, New York, New York (Metropolitan New York City and some neighboring cities, excluding Long Island) (Evolved from R & D center)

Central Atlantic Regional Educational Laboratory, Inc., Alexandria, Virginia (Washington, D.C., and parts of Maryland, Virginia, Delaware, and West Virginia)

Central Midwestern Regional Educational Laboratory, Inc., St. Ann, Missouri (eastern Missouri, southern Illinois, and western Tennessee and Kentucky)

Cooperative Educational Research Laboratory, Inc., Winnetka, Illinois (Indiana, and parts of Illinois, Michigan, and Wisconsin)

Eastern Regional Institute for Education, Syracuse, New York (western Pennsylvania and New York State, excluding New York City)

Far West Regional Educational Laboratory, San Francisco, California (northern California, all of Nevada except Clark County)

Institute for Educational Innovation, Newton, Massachusetts (New England)

Michigan-Ohio Regional Educational Laboratory, Detroit, Michigan (Michigan and Ohio)

Mid-Continent Regional Educational Laboratory, Inc., Kansas City, Missouri (western Missouri, central Oklahoma, and parts of Kansas and Nebraska)

Northwest Regional Educational Laboratory, Portland, Oregon (Alaska, Montana, Oregon, Washington, and northern Idaho)

Regional Educational Laboratory for the Carolinas and Virginia, Rougemont, North Carolina (North Carolina, South Carolina, and southern Virginia)

Research for Better Schools, Inc., Philadelphia, Pennsylvania (eastern Pennsylvania, much of New Jersey and Delaware)

Rocky Mountain Regional Educational Laboratory, Denver, Colorado (all or portions of Colorado, Utah, Wyoming, Idaho, Montana, Arizona, Kansas, and Nebraska)

South Central Region Educational Laboratory Corporation, Little Rock, Arkansas (Arkansas, Mississippi, and portions of Louisiana, Oklahoma, Kansas, and Missouri)

Southeastern Educational Laboratory, Atlanta, Georgia (Florida, Georgia, and Alabama)

Southwest Educational Development Corporation, Austin, Texas (eastern and central Texas and southern Louisiana)

Southwest Regional Educational Laboratory, Englewood, California (southern California, southern Nevada, and southwestern Arizona)

Southwestern Cooperative Educational Laboratory, Albuquerque, New Mexico (all of New Mexico and portions of Arizona, Texas, and Oklahoma)

Upper Midwest Regional Educational Laboratory, Inc., St. Paul, Minnesota (Iowa, Minnesota, North Dakota, South Dakota, and portions of Wisconsin)

APPENDIX F. ERIC CLEARINGHOUSES

During fiscal 1966, the first 13 of a network of ERIC clearinghouses were established. They will provide information on the subject areas listed below.

ERIC Clearinghouse on *Counseling and Guidance*, University of Michigan, Ann Arbor, Michigan 48104

ERIC Clearinghouse on the *Disadvantaged*, Yeshiva University, 55 Fifth Avenue, New York, New York 10003

ERIC Clearinghouse on *Educational Administration*, University of Oregon, Eugene, Oregon 97403

ERIC Clearinghouse on *Exceptional Children*, Council for Exceptional Children, National Education Association, 1201 16th Street, N.W., Washington, D.C. 20036

ERIC Clearinghouse on the Teaching of *Foreign Languages*, Modern Language Association of America, 4 Washington Place, New York, New York 10003

ERIC Clearinghouse on *Junior Colleges*, University of California, 405 Hilgard Ave., Los Angeles, California 90024

ERIC Clearinghouse on *Linguistics* and the Uncommonly Taught Languages, *Center for Applied* Linguistics, 1755 Massachusetts Avenue, N.W., Washington, D.C. 20036

ERIC Clearinghouse on *Reading*, Indiana University, 204 Pine Hall, Bloomington, Indiana 47401

ERIC Clearinghouse on *School Personnel*, City University of New York, 33 West 42d Street, New York, New York 10036

ERIC Clearinghouse on *Science Education*, Ohio State University, 1314 Kinnear Rd., Columbus, Ohio 43212

ERIC Clearinghouse on *Small Schools* and Rural Compensatory Education, New Mexico State University, University Park, New Mexico 88070

ERIC Clearinghouse on *Vocational and Technical Education*, Ohio State University, 980 Kinnear Rd., Columbus, Ohio 43212

Library for *Adult and Continuing Education*, Syracuse University, 107 Roney Lane, Syracuse, New York 13110

APPENDIX G. INSTRUCTIONAL MATERIAL CENTERS FOR HANDICAPPED CHILDREN AND YOUTH

As of the end of fiscal 1966, Instructional Material Centers had been established at the following locations to serve the educational needs of handicapped children and youth.

American Printing House for the Blind, 1839 Frankfort Avenue, Louisville, Kentucky

Colorado State College, Greeley

Department of Special Education, Superintendent of Public Instruction, 302 State Office Building, Springfield, Illinois

Michigan State University, East Lansing

The University of Kentucky, Lexington

The University of Oregon, Eugene

The University of South Florida, Tampa

The University of Southern California, Los Angeles

The University of Texas, Austin

The University of Wisconsin, Madison

U.S. GOVERNMENT PRINTING OFFICE : 1967—O-258-410

SMALL PROJECT RESEARCH

REGIONAL RESEARCH PROGRAM

BUREAU OF RESEARCH

U.S. OFFICE OF EDUCATION

Preliminary Draft

March 1967

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

WHERE TO APPLY

Proposals From	Send To
DISTRICT OF COLUMBIA KENTUCKY MARYLAND NORTH CAROLINA PUERTO RICO VIRGINIA VIRGIN ISLANDS WEST VIRGINIA	Regional Research Program U.S. Office of Education/DHEW Regional Office III 220 Seventh Street, NE. Charlottesville, Virginia 22901
ALABAMA FLORIDA GEORGIA MISSISSIPPI SOUTH CAROLINA TENNESSEE	Regional Research Program U.S. Office of Education/DHEW Regional Office IV 50 Seventh Street NE. Rm. 404 Atlanta, Georgia 30323
ILLINOIS INDIANA MICHIGAN OHIO WISCONSIN	Regional Research Program U.S. Office of Education/DHEW Regional Office V 433 W. Van Buren Street Chicago, Illinois 60607
ARKANSAS LOUISIANA NEW MEXICO OKLAHOMA TEXAS	Regional Research Program U.S. Office of Education/DHEW Regional Office VII 1114 Commerce Street Dallas, Texas 75222
ALASKA ARIZONA CALIFORNIA GUAM HAWAII NEVADA OREGON WASHINGTON AMERICAN SAMOA	Regional Research Program U.S. Office of Education/DHEW Regional Office IX Federal Office Building, Rm. 232 50 Fulton Street San Francisco, California 94102

NOTE: Until educational advisers are appointed for other regions, proposals from all other States should be sent to:

Research Analysis and Allocation Staff
Bureau of Research
U.S. Office of Education
Washington, D.C. 20202

SMALL PROJECT RESEARCH

Regional Research Program

Bureau of Research

Introduction

The Office of Education's Bureau of Research seeks to improve education through support of a variety of educational research and related activities conducted outside the Office with authorizations from the Congress. Supported activities range from small projects (under $10,000) to such comprehensive undertakings as the regional laboratories.* The special provision for Small Project Research is designed to extend opportunities for participation in systematic educational improvement to a greater number of individuals and institutions.

Beginning in the summer of 1966, the administrative responsibility for Small Project Research was decentralized to the Office of Education's Regional Offices, where small project proposals will be received and processed. A list of States served by these offices is shown on page one. This decentralization provides equitable geographic distribution of funds allocated to small projects and facilitates the evaluation of proposed activities in terms of regional needs and other regional educational improvement efforts.

What Small Project Research Is

Definition

To qualify as a Small Project, a proposal for educational research or related activity must meet two basic requirements:

1. The total investment by the Office of Education is $10,000 or less.

2. The project is scheduled for completion within a period of 18 months.

Proposals requesting more support or requiring more time should be submitted through other Bureau programs.

Purpose

The specific purpose of this program is to be more responsive to the regional needs of the educational community by:

1. Providing an opportunity for small projects to receive prompt consideration.

---

* Those interested in other kinds of Bureau-supported programs or projects should consult the appropriate bulletins available from the Bureau of Research.

2. Encouraging small colleges to undertake research programs so that their personnel may gain experience in research and related activities.

3. Supporting significant, small-scale educational research projects by doctoral and post-doctoral students and fellows, particularly those in developing institutions. Funding is out of whichever legislative authorization is most appropriate for the particular kind of activity in question. The determination of this is made by the Bureau of Research staff.

How the Small Project Program Operates

The over-all application instructions of the Bureau of Research (Appendix A) are generally applicable to requests for support of small projects. For example, eligible parties for small projects are the same as for other kinds of activities supported through the Bureau, and the provision for progress, interim, and final reports applies also to small projects.

However, because the general instructions were designed to be uniformly applicable to all kinds of project and program support administered by the Bureau, the applicant for small project support may find the following items helpful in adapting the over-all instructions to accommodate the size and focus of the particular activity he proposes to undertake.

Activities Appropriate for Small Project Support

The criteria for determining whether or not an activity is appropriate for support by Bureau authorizations are the same, regardless of the size of the project. However, as a matter of policy, small project funds may not be used primarily for conducting meetings, conferences, and seminars.

Criteria for Evaluation of Small Project Proposals

Criteria for evaluation of small project applications are the same as those for other projects and programs administered by the Bureau, except that competition is on a regional rather than National basis. In any case, the primary criterion is educational significance--a factor which involves the project's probable impact and its capacity for continuous and effective contribution to educational improvement.

Review Procedures for Small Project Proposals

Small project proposals should be sent to appropriate regional offices (see page 1) where they are assigned OE identification numbers and processed in much the same manner as other proposals are handled in the Bureau. Following staff review, proposals which show promise of contributing to educational improvement are reviewed by specialists outside the Government. Recommendations then go to the Regional Assistant Commissioner as a basis for his decision. Processing from receipt of proposal to notification of action, is expected to be completed within two months, except in rare

instances when complications arise beyond the control of the regional office. The Office does not release information about the status of a proposal until the final disposition is determined. At that time, official notification of the action to be taken is sent to the prospective contractor or grantee, to the initiator, and to the principal investigator. Notification of approval includes the name of the OE staff member who will provide liaison between the Office and the project staff.

Contractual Procedures for Small Projects

The general contractual procedures for other activities receiving support through the Bureau of Research are also applicable to small projects. However, arrangements to handle negotiation and other fiscal matters through the Regional Offices are expected to simplify this process.

Application Procedures

Application for support of a small project is made through submission of a formal proposal document which follows the standard format outlined in Appendix A. However, the small project proposal is generally shorter; fewer copies (20) with one copy of the abstract placed after each title page are needed; and processing can usually be completed in a relatively shorter time than is necessary for larger and more comprehensive activities.

The proposal document is a communication instrument which will be evaluated by the Office of Education staff and field readers selected for their research and other experience and general knowledge of the field. It will be the only contact the reviewers will have with the initiator and his ideas. The proposal must stand on its own merits; if it does not convey the message, staff and field readers will not assume meaning or intent.

It is suggested that the initiator have a person at his institution who is not close to the problem read the proposal to make certain it communicates clearly. Many potentially good research proposals fall short of approval because they are not clearly written.

The following suggestions may be useful to the potential investigator in developing the research proposal:

1. Tie the relevant literature to the objectives of the proposed study. The literature does not have to be exhaustive, but should contain the most pertinent related studies and show an awareness of promising current practices.

2. Do not attempt to attack a problem in a global fashion. In other words, do not present so many questions for research that the study would be unwieldy or could not be completed in a rigorous fashion within the time or funds requested.

3. Use direct language. Present the proposal in a forward manner and be sure that the ideas are clearly delineated. Avoid the use of educational jargon.

4. If the project involves the application of strategy, describe the population characteristics, how they will be selected, the experiences to which they will be exposed, and how the evaluation will be used to determine the effectiveness of the experimental treatment(s). If the activity takes some other approach, be sure the description of activities and the statement of expected outcomes are clear.

Above all, adapt the format to the activity, not the other way around. The suggested format in Appendix A is a guide, not a prescription.

APPENDIX A -- GENERAL APPLICATION INSTRUCTIONS

Activities Appropriate for Support

To be eligible for support administered by the Bureau of Research, an activity must (1) be research or research-related, (2) show promise of improving education, (3) have general (not purely local) applicability, and (4) be directed toward communicable results.

The term "research-related" is broadly interpreted to include such activities as the development of materials and improvement of instructional practices in general and specific areas, and dissemination and implementation of the results of research. Local projects must lead to findings significant for other settings if they are to be considered eligible for support under research authorizations. This applicability or transferability to other settings is the factor which, in many instances, determines whether a given proposal is eligible for consideration by the Bureau of Research.

The Bureau's authorizations do not include funds for general operating programs or for activities carried on in conjunction with the research which are part of the normal activities of project personnel (e.g., teacher salaries in experimental classes). Bureau support may be requested for the research component of larger projects where the operating costs are funded from other sources. In a few instances, support may be provided for those aspects of an operational program which must be modified or established in the interests of the project.

With the exception of Small Project Research and Research Development Grant activities, there are no established time or cost limitations for projects which may be considered for support. However, contractual arrangements usually obligate project funds only out of the current year's appropriations. If a proposed activity involves extensive costs or must be continued over subsequent years, the proposal must undergo substantial additional evaluation to assure sound initial investment.

Eligible Parties

A record of noncompliance with the terms of prior research grants or contracts (for example--failure to submit a final report) by an applicant may be sufficient grounds to refuse consideration of a proposal.

Grants or contracts may be awarded to colleges, universities, State departments of education, or to other public or private agencies, organizations, groups, or individuals after proposals have been reviewed by Office of Education staff and by appropriate non-Government advisory personnel and approved by the U.S. Commissioner of Education. With profit-making organizations, contracts follow Federal Procurement Regulations.

The type of support instrument, i.e., grant, cost-reimbursable contract, or fixed price contract, may be selected by contract personnel during negotiation. If the authorization providing the funds does not specify which type of support, consideration is given to providing the best balance between fiscal control and flexibility of operation. All research grants must include significant cost sharing by the grantee institution. Moreover, in some contract research activities, non-Federal contributions are appropriate and a factor in award consideration.

Eligibility for larger projects and for the beginning phase of longer projects may be limited to groups which are or can be organized so as to ensure proper continued professional and fiscal accountability and competence.

Public schools, school districts, and local educational agencies, in order to assure appropriate liaison and communication with operating school systems, are requested to send three copies of their proposals to their State department of education for comment. To be eligible for support under the Vocational Education Act of 1963, proposals submitted by local educational agencies must contain evidence of approval by an authorized official of the State board of vocational education.

<u>Criteria for Evaluation of Proposals</u>

Proposals are evaluated and recommended for approval according to these criteria:

1. Educational significance.
2. Soundness of design, procedure, or operational plan.
3. Adequacy of personnel and facilities.
4. Economic efficiency.
5. Other specific criteria, as appropriate.

Evaluation of a proposal's significance to education requires more than consideration of the project itself. It involves attention to the breadth of the project's probable impact, its relationship to other ongoing and completed research, and its capacity for contributing to educational improvement within the context of total research needs.

Educational significance, then, is the first of several conditions for support; but if a proposal lacks adequate technical quality, personnel, or economic efficiency, it will not be supported, regardless of its significance. In like manner, no matter how technically excellent the proposal may be, its support is contingent on its significance to education nationally.

Funds are not available to support all the good proposals submitted to the Bureau of Research; the current approval rate is about one in five. To provide a balanced effort toward educational improvement within the context of active competition for available Federal research funds, it is sometimes necessary to forego support in a well-researched area in order to attend to a neglected but critical one. There also are instances when difficult choices must be made between one or two relatively large

projects and a larger number of smaller ones. In any case, selection of any given activity for support is based on systematic evaluation of the plan set forth in the formal proposal document and determination of whether it meets the needs of education.

The Proposal Document

Application for support is made by submitting a properly executed standard application form (available from the Bureau of Research or Regional Offices) and the appropriate number of copies of a formal proposal which describes the activity or activities, explains their significance, identifies key personnel, and estimates costs.

Since funds are available only for the most promising activities, proposals must be assessed competitively (see Criteria for Evaluation of proposals). A uniform proposal format has been designed to accommodate all Bureau of Research projects, regardless of the authorizations under which they are funded. Use of a single format gives the applicant freedom to concentrate on the particular activity to be undertaken without having to choose among application patterns.

All proposals must include the standard cover page, a one-page abstract, the body section, and personnel and budget items. Within this framework, each applicant states the case for his activity.

The guidelines below are generally applicable, regardless of the magnitude of the proposed project, the area of investigation, or the authorization under which the project may be funded, if approved. The applicant is expected to make judicious adaptations of this format to accommodate the kind of research or related activity he proposes to undertake.

I. The Cover Page.--Nothing may precede this page. It contains only the following information in the order indicated, and two copies must be signed. (See page 10 of the pamphlet "Office of Education Support for Research and Related Activities" for correct format.)

II. The Abstract occupies a single page, identifies the proposal, and concisely and simply summarizes the contents. To accommodate the various uses made of this page, the abstract must be written in language understandable by an informed layman. One abstract is placed after the title page in each copy of the formal proposal and 20 copies are provided separately. At the top of the abstract page, the following items should appear on separate lines: Title of Project, Principal Investigator, Contracting Agency, Amount of Federal Funds Requested, and Proposed Beginning and Ending dates. The summary portion of the abstract has three parts: A statement of the purposes, objectives, or nature of the project; an indication of expected contribution to education; and a compendium of procedures or description of what is to be done.

III. The Body of the proposal communicates the project director's plan and its probable effectiveness. It should be clear, concise, forthright, and complete, and should be kept within 20 single-spaced, typewritten pages. Proposals for exceptionally extensive or complicated projects may subsequently be developed into larger documents with the advice and cooperation of Office staff. The body of Small Project Research proposals should not exceed 10 pages.

Within the body section, there are normally three parts which should be subtitled as appropriate for the particular kind of research, development, or dissemination activity to be undertaken. These parts answer the questions: Why? What, when, and how? Use of results? Some attention is also given to personnel, facilities, and costs, which are subsequently treated in detail.

A. Problem and Objectives.--The first part tells why the research or related activity should be undertaken. It includes a statement of the problem or purposes, review of literature and related research, concise statement of objectives, or any other information necessary to establish a sound rationale.

B. Description of Activities (Procedures).--This part tells what is to be done, when, and how. It is the basis for determining the degree to which the proposed activity can be expected to accomplish the objectives or satisfy the need set forth in the first part of the proposal body. It delineates procedures, outlines program arrangements, describes materials to be produced, or otherwise explains how the activity leads to results and how evaluation will be accomplished. Allowances for alternatives, if any, should be noted. A time schedule for completion of the project is usually provided near the end of the proposal.

The amount and kind of detailed information will vary according to the type of activity. For example, a basic research study--which might include rather strict statistical treatment--would be approached quite differently from a curriculum-development activity. A project which is complete within itself would not be approached in the same way as a phase of a proposed multiphase activity. Preparation of initial materials for a modern foreign language would not include the same procedures as the demonstration of new media usage or the testing of a pilot study in vocational or technical education.

C. The Use to be Made of Findings.--This part tells how the results of the activity may be disseminated and/or implemented, what contribution to education can be expected, and what steps should follow.

IV. Personnel and Facilities.--Personnel and facilities are a major determinant of capability.

Personnel with major responsibilities are listed by name, position, title, experience, responsibilities within the project, percentage of time committed to the activity, and the extent to which this commitment has been assured. Consultants who have agreed to serve should be similarly identified (otherwise, the application should describe the type of consultative assistance required).

III. The Body of the proposal communicates the project director's plan and its probable effectiveness. It should be clear, concise, forthright, and complete, and should be kept within 20 single-spaced, typewritten pages. Proposals for exceptionally extensive or complicated projects may subsequently be developed into larger documents with the advice and cooperation of Office staff. The body of Small Project Research proposals should not exceed 10 pages.

Within the body section, there are normally three parts which should be subtitled as appropriate for the particular kind of research, development, or dissemination activity to be undertaken. These parts answer the questions: Why? What, when, and how? Use of results? Some attention is also given to personnel, facilities, and costs, which are subsequently treated in detail.

A. Problem and Objectives.--The first part tells why the research or related activity should be undertaken. It includes a statement of the problem or purposes, review of literature and related research, concise statement of objectives, or any other information necessary to establish a sound rationale.

B. Description of Activities (Procedures).--This part tells what is to be done, when, and how. It is the basis for determining the degree to which the proposed activity can be expected to accomplish the objectives or satisfy the need set forth in the first part of the proposal body. It delineates procedures, outlines program arrangements, describes materials to be produced, or otherwise explains how the activity leads to results and how evaluation will be accomplished. Allowances for alternatives, if any, should be noted. A time schedule for completion of the project is usually provided near the end of the proposal.

The amount and kind of detailed information will vary according to the type of activity. For example, a basic research study--which might include rather strict statistical treatment--would be approached quite differently from a curriculum-development activity. A project which is complete within itself would not be approached in the same way as a phase of a proposed multiphase activity. Preparation of initial materials for a modern foreign language would not include the same procedures as the demonstration of new media usage or the testing of a pilot study in vocational or technical education.

C. The Use to be Made of Findings.--This part tells how the results of the activity may be disseminated and/or implemented, what contribution to education can be expected, and what steps should follow.

IV. Personnel and Facilities.--Personnel and facilities are a major determinant of capability.

Personnel with major responsibilities are listed by name, position, title, experience, responsibilities within the project, percentage of time committed to the activity, and the extent to which this commitment has been assured. Consultants who have agreed to serve should be similarly identified (otherwise, the application should describe the type of consultative assistance required).

Facilities should be described and the extent to which their use has been assured should be indicated.

V. The Budget Section of the proposal starts on a new page and uses the tabular presentation shown on page 15 of the pamphlet "Office of Education Support for Research and Related Activities." The applicant should show reasonable estimates but be detailed enough to suggest careful analysis of expected costs and understanding of fiscal responsibilities in connection with conducting the proposed activity. Alternative budgets may show allowances for extension of objectives or revision of procedures. Local contribution, or cost sharing, is required on all grants, encouraged on contracts. Cost sharing may not be avoided by substitution of a contract in a situation requiring a grant instrument. Local contribution, while not necessarily a measure of the project's soundness or potential worth and not specifically required on certain contracts, is considered evidence of local interest and commitment in the activity. The budget section should include an outline of the sources and amounts of non-Federal support and any conditions upon which this support is contingent.

VI. Appended Items

Other Information.--Give a brief statement about each of the following: (a) If this or a similar proposal has been submitted elsewhere, give details. (b) If this is a proposed extension of, or addition to, a previous or current project supported by the Office of Education, give the Bureau and the grant or contract number of the related support instrument.

Revisions.--If this proposal is a resubmission of a previous formal proposal, give the Bureau proposal number assigned the original proposal and describe the major revisions which have been made. (This does not apply to preliminary statements submitted for informal review.)

Report of Other Projects.--If any of the primary personnel have a current or uncompleted project with the Office of Education or other agency, an appended statement should indicate the status of the project, the amount of time devoted to it, and the relationships between the current and proposed projects.

If any of the personnel have completed a research or development project supported by the Office of Education, give information to identify it. If findings of the previous project are related to the current proposal, summarize them briefly.

Agreement with Cooperating Agencies.--Where agreements with school districts or other cooperating agencies are a factor, copies should be appended.

Instruments.--When an applicant plans to use a published or unpublished instrument (e.g., a questionnaire or interview guide) in his study, the proposal document should include a copy or, if the instrument is still to be developed, a page of sample items and an outline of the complete instrument. (See special conditions, data-collection instruments.)

Other Items.--If necessary, other items may be added, but appended items cannot carry the burden of the request for support.

Details About Submitting Proposals

Forms for the standard application should be requested from the Bureau of Research at least two weeks prior to the date the copies of the formal proposal are to be submitted.

Proposals should be on one side of standard (8½" x 11") paper, stapled at the left margin. Do not bind or enclose in folders. Cover pages on two copies should be signed.

Address.

The address label for most proposals should read:

Research Analysis and Allocation Staff
Bureau of Research
U.S. Office of Education
Washington, D.C. 20202

Note.--If the OE Regional Office in the applicant's area has been staffed with an Educational Research Adviser, proposals for Small Project Research and Research Development Grants should be sent there. Otherwise, they should be mailed to the above address. When in doubt, contact the Bureau or the appropriate Regional Office listed in Appendix B.

Number of Copies.

Small Projects.--Send 20 copies of proposals requesting $10,00 or less in Federal support, with appropriate copies of abstracts.

General Projects.--Send 30 copies of larger proposals. The Bureau may request 10 additional copies of complex proposals. Besides the abstract inside the cover sheet of each proposal, 20 copies of the abstract should be provided separately.

Review Procedures

There are no deadlines for general submission, but some of the special programs process proposals only at certain times. Proposals are first reviewed by Office of Education staff. Those which are within the scope of the Bureau's legislative authority and show promise of contributing to educational improvement are then reviewed by specialists outside the Government. These groups make recommendations to the Commissioner as a basis for his decision and final approval. While many proposals may be processed within three or four months (Small Projects within two months), larger or more complicated ones may require more time.

Communications Regarding Proposals

The initiator will be sent the name of the Bureau of Research unit, the project officer, and the official identification number assigned to his proposal. The Office does not give further information about the status of the proposal during the review and recommendation process. When the proposal has been thoroughly assessed in terms of the criteria listed earlier in this booklet and within the context of pressure for available research funds, the initiator will be notified whether his proposed activity is to be recommended for negotiation. If the project is so recommended, the initiator will be sent any information necessary for subsequent action.

Contractual Procedures

Negotiation is that process by which a legal and mutually satisfactory agreement (contract or grant) is arranged between the applicant and the Office of Education. This process transcends a mere final examination of estimated costs. The contract specialist works with the research specialists (applicants and appropriate staff from the Bureau of Research) to clarify and set forth their mutual agreements about the work and services to be carried out and the support to be provided. The resulting contract or grant instrument, which incorporates the final document and/or plan of operation, thus states the conditions under which the award is made by the Office.

The investigator is cautioned that no costs should be incurred prior to the receipt of a contractual document (grant, contract, or notice of award) executed by the Contracting Officer on behalf of the Government. All contracts are subject to satisfactory fiscal and progress reports. In those cases where only an initial phase of a total research project has been supported, continuation is subject to reappraisal of project aims and accomplishments in relation to emerging research needs and availability of funds in any subsequent period.

Reports

Reports to be received by project officers include: (1) generally brief monthly and/or quarterly progress reports which are intended to facilitate project monitoring; (2) two types of substantive reports, (a) interim reports as appropriate and (b) final reports;* (3) appropriate fiscal reports.

Interim reports include any written material (textual, graphic, or tabular) prepared during the project period and generally before submission of the final report. Often such reports are prepared for delivery at a professional meeting, for distribution to colleagues, or for some other

---

* Detailed instructions for preparing these interim and final reports and related information storage and retrieval items are provided at the time the project is funded.

kind of presentation of substantive findings. Regardless of the number or variety of progress and interim reports submitted, the project director also must submit 15 copies of a comprehensive final report.

For projects that will produce nonprint materials, the project officer should specify at the time of project negotiation the number of copies of such materials to be received by the Office of Education, taking into account loan uses by program staff, possible uses in regional laboratories, and other needs.

Special Conditions

Release of Materials.

Materials produced under research contracts or grants will be placed in the public domain. Whenever any materials not first produced in the performance of a particular contract are to be incorporated in the new materials, the contract must include appropriate arrangements for use of materials so incorporated.

Data-Collection Instruments.

Prior Approval.--Clearance prior to administration is required for all tests, questionnaires, inventories, interview schedules or guides, rating scales, and survey plans used to collect information on identical items from 10 or more individuals or organizations. This regulation flows from the provisions of Federal law, from technical requirements for proper project-monitoring, and from sound considerations of policy in the administration of public funds for educational research.

Description.--All projects involving data-gathering instruments require the submission of a statement reporting briefly: the exact title of the instrument; the purpose(s) of the study and the relation of the data-gathering instruments thereto; the nature and size of the sample, including the method(s) of sampling; the locale of the study; provisions for anonymity and confidentiality of response; a brief indication of the nature and extent of statistical analysis of the data; the estimated average time required of the respondent; and a single figure giving the estimated cost of the survey-proper (as distinguished from the total project). Clearance will take all these factors into account; thus, questionnaire-items that are permissible or desirable in one sample might not be so in another.

Clearance.--The clearance process requires six copies of each instrument, with these exceptions: Copies need not be submitted of instruments which deal solely with cognitive functions or technical proficiency (e.g., scholastic aptitude, school achievement, vocational proficiency); routine demographic information (e.g., age, sex, race, residence, school attendance);

or routine institutional information. However, the information specified in the preceding paragraph is required for all instruments, whether copies of the instrument are submitted or not.

In the process of clearance, attention will be given especially but not exclusively to such matters as the unnecessary or offensive intrusion of inquiries regarding religion, sex, and politics; the extraction of self-demeaning or self-incriminating disclosures; and the apparent countenancing of antisocial or immoral behavior. When the respondents include students below college age, the assurance of informed parental consent will be required if unduly sensitive questions are involved.

Those who will need to clear data-gathering instruments should request separate, more detailed instructions from the Bureau.

Treatment of Animals

If animals are involved in research activities supported by the Bureau of Research, their care and treatment must conform to the principles set forth by the Institute of Laboratory Animal Resources of the National Academy of Sciences--National Research Council.

Research Outside the United States

There are special requirements applicable to research activities outside the United States, including Department of State approval.

APPENDIX B--OE REGIONAL OFFICES

Regional Office of Education Staff are located at the following addresses:

Region	States Served	Address
Region I	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont	Office of Education/DHEW John Fitzgerald Kennedy Federal Building Boston, Massachusetts 02203
Region II	Delaware, New Jersey, New York, Pennsylvania	Office of Education/DHEW 42 Broadway New York, New York 10004
Region III	Kentucky, Maryland, North Carolina, Puerto Rico, Virginia, Virgin Islands, West Virginia, District of Columbia	Office of Education/DHEW 220 Seventh Street, N.E. Charlottesville, Virginia 22901
Region IV	Alabama, Florida, Georgia, Mississippi, South Carolina, Tennessee	Office of Education/DHEW 50 Seventh Street, N.E. Atlanta, Georgia 30323
Region V	Illinois, Indiana, Michigan, Ohio, Wisconsin	Office of Education/DHEW New Post Office Building 433 West Van Buren Street Chicago, Illinois 60607
Region VI	Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota	Office of Education/DHEW 601 East 12th Street Kansas City, Missouri 64106
Region VII	Arkansas, Louisiana, New Mexico, Oklahoma, Texas	Office of Education/DHEW 1114 Commerce Street Dallas, Texas 75222
Region VIII	Colorado, Idaho, Montana, Utah, Wyoming	Office of Education/DHEW Federal Office Building, Rm. 9017 19th and Stout Streets Denver, Colorado 80202
Region IX	Alaska, Arizona, California, Guam, Hawaii, Nevada, Oregon, Washington, American Samoa	Office of Education/DHEW Federal Office Building, Rm. 232 50 Fulton Street San Francisco, California 94102